

THE DEVELOPMENT OF THE OBLIGATION SYSTEM IN MEXICAN SPANISH: A  
VARIATIONIST SOCIOLINGUISTIC PERSPECTIVE

By

KATHERINE HONEA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

© 2012 Katherine Honea

To my sister, the memory of your laugh, your smile, or those nights of us just laying on the bed talking has helped get me through the most stressful of times, I love you

## ACKNOWLEDGMENTS

I want to express my gratitude to a number of people who have supported me throughout my education and through this dissertation process. First, I want to thank the faculty for all their support. I am eternally indebted to Jessi Aaron, who has spent countless hours offering invaluable advice, both as an advisor and as a friend. I remember so clearly that first class I took and the overwhelming yet satisfying feeling I had as she opened my eyes to functionalism. That class was a turning point for me and I am truly grateful for that. The opportunity to work with such a remarkable scholar and to learn so much about functionalism and language change has been an experience that I will never forget. I am also grateful to Ana de Prada Perez and Geraldine Nichols for the many lessons they have taught me about writing and organization. It is because of them that my writing has improved so much over the last four years. Many thanks also go to Gillian Lord, who has been supportive since the day I arrived in Gainesville. She has believed in me and given me the opportunity to teach content courses and has always encouraged me incorporate new ideas into my lessons. She has been a tremendous support throughout my career and I feel lucky to have her as a friend and colleague. I am equally appreciative to David Pharies and Diana Boxer for being the voices of reason throughout this process. Whether it was through occasional visits to their offices, or through emails, they have always managed to keep me grounded or make me laugh. I am grateful to Jason Rothman for being so supportive of my development as a professor. Whether it was his advice in my practice job talks or conference presentations or his supervision and training of my teaching, he has helped build my confidence and has truly prepared me to teach content-based courses. I would

also like to thank Wind Cowles, who allowed me to look at this research project from a new perspective with her comments and suggestions.

In general, the Spanish and Portuguese Studies department has provided me with endless support, ranging from Bobbye and Tania making sure all the necessary paperwork was done, to mock interviews and job talks with half of the faculty present. I am reminded on a daily basis of how lucky I am to have been part of such a supportive department.

Of course, I cannot begin to say how grateful I am to my friends and family. The countless ways they have shown their support through letters, phone calls or just listening to me have been a guiding light throughout my life. Knowing that I have support from friends and family like Diego, Laurie, Briane, Valerie, Donna, Don, Raciél, Diana, Kathleen, Sabrina, Brian, Jake, Shawna, Eric, Bill, Sandy, Pilla, Ariana and all my friends in Guanajuato (and also those who I have not mentioned) has been a constant comfort and driving force in my life. I have always known that I am surrounded by some exceptional people, but this process has made that fact even more apparent to me. I am particularly grateful to Delano Lamy, Maria Fionda and Kimberly Louie who have been pillars of support for me. I would not be here today if I didn't have each of you in my life. You three have listened to and sympathized with every painful (and joyful) step in this process. I love you all so very much and am fortunate to count you three as part of my extended family.

A special thank you is in order for my parents, who have full-heartedly supported me in every adventure I have ever embarked on in life, despite their own personal misgivings. They have somehow always managed to know when I needed a little break,

a quick laugh, a homemade card from my nieces or just a hug. They define true support and I am forever grateful to the two of them.

I would also like to thank my little Gainesville family, my husband Chad and my cats Muki, Chucha, and Marley, who continue to love me unconditionally despite months and months of stress. I am truly blessed to be married to such an accommodating, dedicated, and compassionate man. The unfaltering support he gives me through the countless meals he cooks, the constant words of encouragement, and the weekly serenades continue to keep me sane. Also, I do not know how I would have made it through each day, locked up writing in my house, without the presence of my three cute and devoted kitties.

Ultimately, this research project would not have been possible without the careful guidance of my committee members, the continuous support of my department, and the love and confidence from my friends and family. I have so many people to thank for helping me get through this program.

## TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	10
LIST OF FIGURES.....	12
ABSTRACT .....	13
CHAPTER	
1 INTRODUCTION .....	15
1.1 Defining Modality in a Dynamic System.....	16
1.1.1 Classification Systems and Terminology .....	16
1.1.2 Diachronic Change in Modality.....	23
1.1.3 Modality in Spanish .....	26
1.2 Usage-Based Models.....	29
1.2.1 Grammaticalization Theory.....	31
1.2.2 Mechanisms in Grammaticalization .....	36
1.2.3 Subjectification .....	40
1.2.4 Subjectification and Grammaticalization.....	44
1.3 The Obligation System in Spanish.....	46
1.3.1 The (Non)grammaticalization of These Constructions.....	47
1.3.2 The (Non)subjectification of These Constructions .....	51
1.4 A Variationist Perspective .....	54
1.4.1 The Variationist Method.....	54
1.4.2 Comparative Method .....	56
1.5 The Present Study .....	57
2 METHODOLOGY .....	59
2.1 Corpora.....	59
2.1.1 Written Corpora .....	59
2.1.2 Oral Corpus .....	61
2.1.2.1 The speech community .....	62
2.1.2.2 Participant selection .....	63
2.2 Data Collection .....	65
2.3 The Variable Context and Form-Function Asymmetry .....	66
2.4 Circumscribing the Variable Context.....	67
2.5 Variable Rule Analysis .....	68
2.6 Comparative Sociolinguistic Method.....	71
2.7 Variables and Hypotheses .....	74
2.7.1 Verb Type.....	75
2.7.2 Subject and Animacy.....	76

2.7.3 Sentence Type .....	78
2.7.4 Polarity .....	80
2.7.5 Objects .....	81
2.7.6 Preceding Clitic .....	83
2.7.7 Temporal Adverbs .....	84
2.7.8 Tense, Aspect, Mood.....	86
2.7.9 Lexical Verb Type.....	87
2.7.10 Extralinguistic Variables .....	89
2.7.10.1 Age.....	90
2.7.10.2 Sex.....	91
2.8 Summary .....	92
<b>3 COMPETITION, VARIATION AND LANGUAGE CHANGE .....</b>	<b>93</b>
3.1 Grammaticalization of Changing Forms.....	93
3.2 Frequency and Language Change .....	96
3.2.1 Verb Type.....	98
3.2.1.1 <i>Tener que</i> .....	98
3.2.1.2 <i>Haber de</i> .....	100
3.2.1.3 <i>Haber que</i> .....	101
3.2.1.4 <i>Tener de</i> .....	102
3.2.2 Grammatical Person and Animacy .....	103
3.2.2.1 <i>Haber de</i> .....	105
3.2.2.2 <i>Tener de</i> .....	106
3.2.2.3 <i>Tener que</i> .....	107
3.2.2.4 <i>Haber que</i> .....	108
3.2.3 Sentence Type .....	109
3.2.3.1 <i>Haber de</i> .....	109
3.2.3.2 <i>Tener de</i> .....	110
3.2.3.3 <i>Tener que</i> .....	111
3.2.3.4 <i>Haber que</i> .....	111
3.2.4 Polarity .....	112
3.2.5 Preceding Clitic .....	113
3.2.6 Direct Objects.....	114
3.2.6.1 <i>Haber de</i> .....	115
3.2.6.2 <i>Tener de</i> .....	116
3.2.6.3 <i>Tener que</i> .....	116
3.2.6.4 <i>Haber que</i> .....	117
3.2.7 Indirect Objects .....	118
3.2.8 Tense-Aspect-Mood .....	120
3.2.8.1 <i>Haber de</i> .....	121
3.2.8.2 <i>Tener de</i> .....	122
3.2.8.3 <i>Tener que</i> .....	123
3.2.8.4 <i>Haber que</i> .....	124
3.2.9 Temporal Adverbs .....	124
3.3 Multivariate Analyses.....	126
3.3.1 <i>Haber de</i> vs. <i>Tener que</i> : 19 <sup>th</sup> Century .....	128

3.3.2	<i>Haber de vs. Tener que: 20<sup>th</sup> Century</i> .....	130
3.3.3	<i>Haber de vs. Haber que: 20<sup>th</sup> Century</i> .....	134
3.3.4	<i>Tener que vs. Haber que: 20<sup>th</sup> Century</i> .....	135
3.4	Comparative Analysis .....	137
3.5	DLNE and Oral Corpus .....	139
3.6	Summary .....	144
4	DATA ANALYSIS.....	145
4.1	Identifying the Path of Change.....	147
4.1.1	Grammaticalization.....	148
4.1.2	Subjectification .....	153
4.2	Data Analysis.....	155
4.2.1	Tener de.....	156
4.2.2	Tener que .....	161
4.2.3	Haber de.....	165
4.2.4	Haber que.....	168
4.3	General Discussion.....	172
5	FUTURE DIRECTIONS AND CONCLUSIONS .....	181
5.1	Future Directions.....	181
5.2	General Summary.....	184
APPENDIX		
A	LIST OF TEXTS USED IN WRITTEN CORPORA.....	189
B	SAMPLE QUESTIONS FOR SOCIOLINGUISTIC INTERVIEW .....	195
C	COMPLETE LIST OF FACTOR GROUPS AND FACTORS.....	198
LIST OF REFERENCES .....		200
BIOGRAPHICAL SKETCH.....		212

## LIST OF TABLES

<u>Table</u>	<u>page</u>
1-1 Traugott's (1989) path of subjectification of modals (re-printed with permission) .....	53
2-1 Frequency count of each variant per century in the written corpora .....	60
2-2 Frequency count of each variant per century in the oral and DLNE corpora .....	61
2-3 Exclusions of <i>tener que/de</i> (T) and <i>haber de/que</i> diachronically .....	66
3-1 Absolute and relative frequency of each variant per century in the written corpora normalized per 5000 words .....	97
3-2 Frequency of <i>tener que</i> with the factor group verb type .....	99
3-3 Frequency of <i>haber de</i> with the factor group verb type .....	101
3-4 Relative frequency of <i>haber que</i> with the factor group verb type .....	102
3-5 Frequency of <i>tener de</i> with the factor group Verb type .....	103
3-6 Frequency of <i>haber de</i> with the factor group grammatical person .....	105
3-7 Frequency of <i>tener de</i> with the factor group grammatical person .....	106
3-8 Frequency of <i>tener que</i> with the factor group grammatical person .....	108
3-9 Frequency <i>haber que</i> with the factor group grammatical person .....	109
3-10 Frequency <i>haber de</i> with the factor group type of sentence .....	110
3-11 Frequency <i>tener de</i> with the factor group type of sentence .....	110
3-12 Frequency <i>tener que</i> with the factor group type of sentence .....	111
3-13 Frequency <i>haber que</i> with the factor group type of sentence .....	111
3-14 Frequency of <i>all constructions</i> with the factor group polarity .....	112
3-15 Frequency of <i>all constructions</i> with the factor group preceding clitic .....	113
3-16 Frequency of <i>haber de</i> with the factor group direct objects .....	115
3-17 Frequency of <i>tener de</i> with the factor group direct objects .....	116
3-18 Frequency of <i>tener que</i> with the factor group direct objects .....	117

3-19	Frequency of <i>haber que</i> with the factor group direct objects .....	118
3-20	Frequency of all factors occurring with the factor group indirect objects .....	119
3-21	Frequency of <i>haber de</i> occurring with the factor group tense-aspect-mood.....	122
3-22	Frequency of <i>tener que</i> occurring with the factor group tense-aspect-mood....	123
3-23	The occurrence of all obligation markers with temporal adverbial Markers across centuries .....	125
3-24	Analysis from the 19th century: Linguistic factors conditioning <i>haber de</i> occurrence in comparison with <i>tener que</i> in CORDE. ....	130
3-25	Data from the 20th Century: Linguistic factors conditioning <i>haber de</i> occurrence in comparison with <i>tener que</i> in CORDE. ....	133
3-26	Data from the 20 <sup>th</sup> century: Linguistic factors conditioning <i>haber de</i> occurrence in comparison with <i>haber que</i> in CORDE .....	135
3-27	Data from the 20 <sup>th</sup> century: Linguistic factors conditioning <i>tener que</i> occurrence in comparison with <i>haber que</i> in CORDE. ....	136
3-28	Variation of <i>haber de</i> and <i>tener que</i> across two centuries.....	137
3-29	Analysis of constrain hierarchies of <i>haber de</i> vs. <i>tener que</i> across the 19th and 20th centuries. ....	138
3-30	Data from the 21 <sup>st</sup> Century: Frequency of occurrence of each obligation marker within factor group verb type .....	140
3-31	Data from the 21 <sup>st</sup> Century: Frequency of occurrence of each obligation marker according to factor group.....	142
3-32	Data from the 21 <sup>st</sup> Century: Frequency of occurrence of each obligation marker within factor group tense-aspect-mood (TAM).....	143

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
4-1 Frequency of variants across centuries .....	146
4-2 Comparison of <i>tener de</i> as an possession marker and as an obligation marker across centuries .....	161

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

THE DEVELOPMENT OF THE OBLIGATION SYSTEM IN MEXICAN SPANISH: A  
VARIATIONIST SOCIOLINGUISTIC PERSPECTIVE

By

Katherine Honea

August 2012

Chair: Jessi Aaron  
Major: Romance Languages

Research that examines diachronic change and modality posit that modal verbs follow certain universal paths of development (e.g. Cornillie, 2007; Bybee & Fleischman, 1995; Bybee, Perkins & Pagliuca, 1994). The present study examines the development of Spanish modality in Mexico through the use of multivariate analyses, relative frequencies, and the comparative method as a means to uncover where in the grammar changes are taking place. In addition, this study exposes the subtle semantic differences of four obligation markers – *haber de*, *tener de*, *tener que*, *hay que* – commonly used in Mexico between the 16<sup>th</sup> and 21<sup>st</sup> centuries.

Results from this study aim to explain the use of one marker over another by operationalizing various factors found to contribute to changes in modal systems across languages. The factors examined include: type of verb, animacy/grammatical person, type of sentence, tense-aspect-mood, temporal markers, polarity, objects (direct, indirect), preceding clitic, lexical verb type, sex and age. Working within the usage-based framework, 5691 tokens were extracted, analyzed and compared across centuries using relative frequencies and the statistical program GoldVarb (Cedergren and Sankoff, 1974).

Results indicated that there is an overall shift in preference from *haber de* to *tener que*, suggesting that the system is undergoing longitudinal renewal. This is supported by the dramatic increase in frequency of *tener que* between the 19<sup>th</sup> and 20<sup>th</sup> centuries and the sharp decrease of *haber de* during this same period. A statistical examination of this shift indicates that change is occurring within grammatical person/animacy. Research on modality (e.g. Bybee and Fleischman, 1995) suggests that, as obligation markers develop, they will occur more often in contexts with third-person subjects so it is not surprising that animacy/grammatical person may be an area in the grammar where change takes place. In addition to the outcome from the 19<sup>th</sup> and 20<sup>th</sup> centuries, the results also expose the various factors that condition the use of each of these obligation markers over time. Through evidence found in processes such as semantic weakening, decategorialization, the results corroborate postulations of a universal path of change in modality.

## CHAPTER 1 INTRODUCTION

Modality is an essential element of human experience and thus a fundamental part of how we think and express ourselves. It comprises possible and necessary truths and allows us to form thoughts and/or statements about whether something is conceptually, logically, physically or metaphysically possible or necessary (Melia, 2003). These truths are realized in language through numerous channels. For example, in Spanish modality is expressed in a variety of ways, from mood –subjunctive, indicative- (e.g. Solano-Araya, 1984; Silva-Corvalán, 1985), to modal adverbs –*probablemente* ‘probably’, *ciertamente* ‘certainly’- (e.g. Hengeveld, 1988), to lexical verbs –*obligar* ‘oblige’, *necesitar* ‘need’- (e.g. Palmer, 2001), or modal periphrases (e.g. Yllera, 1980). It is the expression of modality through the use of modal periphrases that is the focus of the present study, particularly *tener de* (1.1), *hay que* (1.2), *tener que* (1.3), and *haber de* (1.4) all of which express strong obligation.

- 1.1 *Fabio Yo **tengo de** ir con vos.* (Pedro Calderón de la Barca, 1629: CORDE)  
‘Fabio I have to go with you.’<sup>1</sup>
- 1.2 *No **hay que** mirarme, que vos no merecisteis descalzarme.* (Moreto, Agustín, 1658: CORDE)  
‘You don’t have to look at me, you didn’t deserve to take my shoes off.’
- 1.3 *Siendo esa vuestra opinión, ya no **tengo que** os decir.* (Calderón de la Barca, Pedro, 1649)  
‘This being your (pl) opinión, I don’t have to tell you (pl) anymore.’
- 1.4 *Señor, ¿con qué mandamiento te **he de** matar?* (Quirós, Francisco Bernardo de, 1656: CORDE)  
‘Sir, with what order do I have to kill you?’

---

<sup>1</sup> All translations are mine unless otherwise noted.

The purpose of this study is to examine these strong obligation markers diachronically. This study will contribute to the limited literature on diachronic variation and change within modality in Spanish.

### **1.1 Defining Modality in a Dynamic System**

The concept of modality has long been a topic in the field of logic. Outside of linguistics, it has traditionally been associated with possibility and necessity (Coates, 1983; Kiefer, 1987; Narrog, 2005a, 2005b). Unsurprisingly, this association has led to a tendency to classify linguistic modality in a similar way, which has been problematic at best, as we will see below in section 1.2.1. Subsequently, a large body of research on modality has focused primarily on two areas: 1) defining and classifying modality at both the macro and the micro level; and 2) diachronic changes that occur in modal systems. In the following sections, these areas will be examined.

#### **1.1.1 Classification Systems and Terminology**

Defining modality is a particularly challenging task, given that it is a universal feature of human languages and therefore categorizing it should necessarily encompass modality in all languages (Coates, 1983; Bybee, Perkins, & Pagliuca, 1994; Narrog, 2005a). Although there is general agreement that modality is a semantic category (Fleischman, 1982; Bybee & Pagliuca, 1985; Bybee & Fleischman, 1995; Palmer, 2001), when it comes to a formal definition, researchers have disagreed as often as they have agreed. Palmer (2001: 1) for example, states that modality refers specifically to the proposition that describes an event, while Kiefer (1987) views it as related to possibility and necessity, basing his conceptualization on modal logic. On the other hand, Bybee et al. (1994: 176) suggest it may be impossible to define modality in such a way that includes both 'a notional domain of modality and the part of it that is

expressed grammatically.’ They therefore take a different tactic and avoid a definition altogether, instead opting to let the results of a cross-linguistic diachronic study on modality speak for themselves. Coates (1983: 10) also voices her concerns about achieving a uniform definition, stating that, ‘formalisations of modality [...] are concerned with precise categories and have failed to acknowledge the problem of indeterminacy in natural language.’ However, Coates (1983: 235) goes on to broadly categorize modality as ‘non-factive.’ Other linguists have echoed Coates definition (e.g. Lyons, 1977; Narrog, 2005a, 2005b) and its proclivity to encompass the many types of modality. Thus, Coates’ definition is the basis of the definition I will use here, but as put forth by Narrog (2005a, 2005b), who uses the term FACTUALITY:

Modality is a linguistic category referring to the factual status of a state of affairs. The expression of a state of affairs is modalized if it is marked for being undetermined with respect to its factual status, i.e. is neither positively nor negatively factual. (p. 679)

Taking this definition into account, and everything factuality can include in a language, it is not surprising that linguists have deemed it necessary to establish a typology of the different categories of modality that the various levels of factuality include. It is widely accepted that modality is comprised of mood and modality systems; however, there is no consensus on how to separate these into more precise categories. Heine, Claudi and Hünemeyer (1991) reiterate the ambiguity of modality and the general difficulty of defining forms by discrete categories. They claim that the use of discrete categories does not give a complete picture of the form because, in reality, the meaning of a form is part of a continuum, often referred to in the grammaticalization literature as a cline, and can vary greatly:

This means that, rather than analyzing the structure...in terms of discrete categories such as constituent types or morpheme classes, a more

appropriate approach would be that which highlights the continuum nature of linguistic structures (1991: 3).

This approach, however, has not been widely adopted in the literature. Instead, a wide array of classifications has been postulated to account for its semantic (and grammatical) multiplicity. Generally, these typologies fall into one of two categories: bipartite systems, such as those of Lyons (1977), Givon (1990 [1984]), Bybee and Fleischman (1995), and Palmer (2001); and those systems that consist of more than two categories, like those proposed by Bybee et al. (1994), Skotarek, (1996) and Narrog (2005a, 2005b).

Of those that have a binary classification, the most traditional typology is DEONTIC/EPISTEMIC (e.g. Solano-Araya, 1982; Brinton, 1991), where deontic 'relates to obligation or permission, emanating from an external source' (Palmer, 2001: 10) and epistemic is 'concerned with the speaker's assumptions or assessment of [...] the truth of the proposition' (Coates, 1983: 18). Some authors, however, have found this typology to be too restrictive. Palmer (2001), for example, also divides modality into two systems, EVENT MODALITY and PROPOSITIONAL MODALITY, but differentiates his classification by asserting that his taxonomy is more general and thus covers a broader range of modality. He cites DYNAMIC MODALITY –related to ability or willingness– as an example of event modality that would otherwise be excluded from the traditional deontic/epistemic typology. Coates (1983) also addresses this issue by using the term ROOT MODALITY in place of deontic modality in order to cover a larger subset of modals (essentially adding ability to the group).

Bybee and Pagliuca (1984) build on Coates' (1983) definition of root modality in their study on modals but use the term AGENT-ORIENTED MODALITY. They attempt to

broaden the scope of classification by positing a binary system of agent-oriented modality and epistemic modality. They define agent-oriented modality as ‘modalities that predicate conditions of either an internal or external nature on a willful agent,’ which essentially includes ability, obligation, desire and intention (p. 63). Again, like Palmer (2001) and Coates (1983), the aim was to include a broader range of modals than the prototypical deontic/epistemic typology.<sup>2</sup> All of these linguists address a valid issue by modifying the traditional deontic/epistemic classification while maintaining the bipartite system.

In addition to binary classifications there have also been typologies including more than two categories. The purpose of these typologies has been to not only account for the various types of modality not included in the aforementioned binary classifications, but also to account for typical changes found in modal systems, which will be discussed in further detail in section 1.2.2.

Narrog (2005a) examines modality from a dimensional perspective by postulating two modal axes. On one end of the first axis are modals that are more objective, or EVENT-ORIENTED (e.g. grammatical third-person would be more event oriented) and on the other end are those that are more subjective or SPEAKER-ORIENTED (e.g. grammatical first-person would be more speaker oriented). She calls the other axis the VOLITIVITY dimension, which on the volitive end she equates to deontic modality and the other end, non-volitive, she correlates this end to epistemic, evidential and dynamic modality,. In essence, on one end of this dimension are modals that require some ‘action that has to

---

<sup>2</sup> As Narrog accurately points out the term ‘agent-oriented’ can lead to confusion since in many examples the subject is not always the agent but can also be an experiencer or even not expressed (2005a: 682). Nevertheless, for the present paper this term will be adopted since it is part of the terminology in the theory of change proposed by Bybee et al. (1994) that will be used throughout this paper.

be performed volitionally by the addressee or some other person understood in the context' (2005a: 683). On the other end are modals that don't require such an action. While her argument for defining changes in modality with volition is convincing and her attempt to include SUBJECTIFICATION (i.e. speaker-oriented) as a principle dimension is valuable, her model only provides a general path of development for all modals. In the classification put forth by Bybee et al. (1994), on the other hand, they identify the paths of each of the different types of modals and suggest specific paths (e.g. obligation>intention>future>purpose/to want or order) for each of the different types of modals thereby offering a more specific account of development for modality.

Bybee et al. (1994) identify four categories within modality systems: AGENT-ORIENTED, EPISTEMIC, SPEAKER-ORIENTED and SUBORDINATING MOODS. Similar to Bybee and Pagliuca (1984), they define agent-oriented modality as an element of internal or external will that is placed on the agent. They include in this classification obligation, necessity, desire, and ability (1994: 178). Speaker-oriented modality includes directives and utterances where permission is granted to the listener. They emphasize that speaker-oriented modality does not include conditions (external or internal) on the agent, but instead allows the speaker to enact these conditions (p. 179). Their designation of epistemic modality concurs with other definitions by labeling it as the commitment of the speaker to the truth of the proposition (p. 179). The last modality type, subordinating moods, comprises the same constructions that are used in agent (e.g. obligation) and speaker-oriented modality (e.g. directives) but are markers of the verb in specific types of subordinate clauses such as in the example given by Bybee et al. (1994: 180) in 1.5. This typology is complete in that it incorporates an extensive

range of modality (like illocutionary force, lexical verbs etc.) and it also considers mood in its typology which is often regarded as separate. Consequently, it is the typology adopted for the present dissertation. This does not negate, however, that this typology (as well as the others) fails to address the semantic nuances of these modals on a more micro-level.

1.5 I suggested that he **should** call you immediately.

Determining the semantics of these modal markers has been as polemic as categorizing the different types of modals. Much of the research conducted on the semantic side of modality has determined the meanings of modals based on the psychological motives of the speaker (De Maeseneer, 1998; Myhill, 1995; Sirbu-Dumitrescu, 1988). For example, Myhill (1995) claims that the changes that have occurred in the English modal system are due to an overall shift in societal norms pointing out that, in the last century, the United States has undergone a collective shift from a focus on power and speaker control to a society that is focused on more 'interactive factors' like cooperation, advice etc. He claims that this shift is reflected in the development of the American English obligation system. Although the shift in society is observable, and thus is a possible explanation for changes found in the obligation system, it is difficult to attribute these changes to the societal shift without having a way to operationalize it. Myhill (1996) attempts to do this by explaining that the interest of the speaker is what determines which obligation marker will be used. He categorizes modals with such names as 'benefits relationship between speaker and listener' or 'meeting societal expectations.' A shift based on the motives of the speaker is difficult to prove, however.

In fact, the use of psychological motives as a way to determine semantic categories has been questioned in recent research. This is highlighted in a study on Romance futures by Poplack and Malvar (2007: 162), who state that psychological motives have little empirical basis and that, in order to determine what conditions the use of one variant or another, empirical based analyses are required:

very few of the motivations ascribed to variant choice in the (prescriptive or descriptive) literature are now, or ever have been, relevant to actual usage. Grammarians have been silent on the role of the operative contextual factors, focusing instead on semantic, psychological and other motivations which have no basis in empirical fact.

Other modality studies have focused less on psychological motives, and many of these have determined the semantics of these modals through the use of frequency distributions. Cornillie (2007), for example, uses relative frequency as a measure in his study of modality in Spanish. He looks at the frequency with which *tener que* occurs in certain surrounding contexts revealing specified patterns of use. However, he looks primarily at the patterns of use of the epistemic use of *tener que*.<sup>3</sup> Tagliamonte & D'Arcy (2007) also look at obligation in their study of Toronto English, where they determine the use of the obligation markers by examining the effect that various linguistic and extralinguistic factors have on these constructions. Studies such as these provide an objective means of determining the semantics of these modals and are preferable because they allow for easier replication.

As mentioned above, Heine et al. (1991) suggest that one of the possible reasons for the challenge in uncovering the specific functions of these modals is because

---

<sup>3</sup> Cornillie reached numerous conclusions that are beyond the scope of this study, since his research was related to modal grounding in relation to evidential and epistemic modality.

sometimes the uses can be ambiguous. Take the following sentence taken from Traugott and Dasher (2002: 2) as an example.

1.6 They **must** be married.

They describe that this sentence, when taken out of context, can have two separate meanings: one expressing obligation and one with an epistemic sense. In the obligation sense it would mean something like 'It is necessary that they get married' whereas in the epistemic sense there is a high sense of certainty, essentially expressing that the speaker is certain that the couple is married. As Traugott and Dasher (2002) point out, modals such as this example are polysemic and occur in ambiguous contexts and this ambiguity must be taken into consideration.

Ultimately, all of these studies reveal that many modals are polysemic (e.g. Coates, 1983) and appear to develop and acquire new meanings as well as lose old meanings over time. One question that is raised by this perspective, then, is precisely how do these modal markers acquire multiple meanings?

### **1.1.2 Diachronic Change in Modality**

Change has long been recognized as part of language and modality is not spared in this area. In fact, as previously mentioned, many of the typological studies (e.g. Lyons, 1977; Coates, 1983; Bybee et al., 1994; Narrog, 2005a, 2005b) have identified patterns of change in modal systems. With the identification of these patterns, some linguists (c.f. Lyons, 1977, Bybee & Pagliuca, 1985, Traugott 1989) have taken it a step further and identified universal tendencies in modality. That is they postulate that certain types of modality appear to develop in parallel ways across languages. Most notable is the model for universal changes in modality put forth by Bybee et al. (1994), in which they outline several important universal tendencies through their data.

One such proclivity is the frequency with which they find agent-oriented modality (obligation, necessity, desire and ability) to be periphrastic (e.g. *have to*, *tener que*) while the other three types (speaker-oriented, epistemic, and subordinating moods) were more likely to be inflectional (e.g. synthetic future tense in Spanish: *hablaré* 'I will speak').<sup>4</sup> Taking into consideration that many linguists (e.g. Heine et al., 1991; Bybee et al. 1994) suggest that modality is better viewed as a continuum, or what is often referred as a cline '[t]his correlation suggest[s] a diachronic scenario whereby as agent-oriented modalities GRAMMATICALIZE (i.e. change), they develop into the other types and gradually take on inflectional expression' (Bybee et al. 1994: 181, and references therein). This assumption suggests then, that agent-oriented modality in Spanish should be comprised of periphrastics, which we see with the strong obligation markers *tener que*, *tener de*, *haber de*, and *hay que*. Westney (1995) and Coates (1983) go a step further in suggesting that periphrastics that operate as auxiliary verbs are more developed along this cline than those that do not. Westney (1995: 31) claims that these modal auxiliaries are more likely to occur in epistemic contexts and states that the 'periphrastic items [i.e. semi-auxiliaries] tend to manifest epistemic senses only to a relatively restricted degree.'

As a result of the differences in development among semi-auxiliaries and auxiliaries, criteria for determining the status of a modal verb as either auxiliary or periphrastic have been established. Most notable among these are Palmer's (1979) 'modal criteria' (Westney, 1995, and references therein). If a verb does not meet these

---

<sup>4</sup> Bybee et al. (1994: 180) find that, of the 76 languages examined, 75% of the obligation markers were periphrastic and only 25% were affixes. This was a similar trend found in the various agent oriented modals for which they coded

criteria, it is not considered to be a modal auxiliary but rather a periphrastic or lexical verb (Westney, 1995: 15). Considering that these criteria are specific to English, it is questionable whether the criteria Palmer (1979) and others use are valid in Spanish. Additionally, Bybee et al. (1994) do not identify the development of periphrastics into modal auxiliaries as a universal tendency in modality, although they do use the term 'auxiliary' regularly when mentioning agent-oriented modality thus implying that modal verbs can and do become auxiliaries.

Another tendency Bybee et al. (1994) document is a unidirectional path of change from agent-oriented modality to epistemic modality.<sup>5</sup> This path of change, a process referred to as grammaticalization and, in certain stages, SUBJECTIFICATION, will be discussed in greater detail in section 1.3.1. Given that epistemic modals tend to be inflectional, it can be deduced that the continued progress of these agent-oriented modals towards epistemicity is accompanied by phonological reduction or fusion that ultimately results in inflection. The results of Bybee et al. (1994) reveal that these patterns occur when lexical verbs covering a small range of specific meanings begin to generalize semantically and accordingly move into new contexts. Similarly, they find that once epistemicity is part of the semantics of a modal marker, further generalization could occur whereby these forms could develop into subordinating moods.<sup>6</sup> The tendencies observed by Bybee et al. (1994) highlight the universality in modality and the formation of modals and mood systems. Given the relative youth of the field, however,

---

<sup>5</sup> Narrog tries to challenge this idea by positing his two dimensional model. He claims that there is indeed unidirectionality but the path is from objective to subjective, not agent-oriented to epistemic (2005a).

<sup>6</sup> There is another possible path for agent-oriented modality proposed by Bybee et al. whereby these markers acquire a speaker-oriented notion, then generalize to express subordinating mood (1995; 241). Owing to the fact that the obligation markers in this paper do not appear to take this path, it is not the central focus.

the empirical evidence supporting or refuting claims of cross-linguistic diachronic universality is controversial and has prompted much research and heated debate (e.g. Nordlinger & Traugott, 1997; Axelrod, 1999; Kim, 2009). In order to corroborate or refute these claims of universality more empirical evidence is needed. This study is a step in that direction.

### 1.1.3 Modality in Spanish

In Spanish, as in many other languages, modality is comprised of both mood and a modal system and can be expressed through the use of inflection (e.g. *-ra*), periphrastics (e.g. *tener que*, *deber de*), adverbs (e.g. *probablemente*, *posiblemente*) or even lexical verbs (e.g. *mandar*, *sugerir*). Despite the association between mood and modality, the focus here will primarily center on modal markers given that these strong obligation constructions have yet to become inflectional.

Spanish has not escaped the disagreement described above in terms of defining and classifying modality but research documenting modality in Spanish is minimal. Besides research focused on developing a typology for modality across languages — including Spanish—the limited research available on modality in Spanish has centered on the same two areas previously mentioned: the semantics of these individual modal markers and the diachronic changes these modals have undergone.

Molina Plaza (2005) for example looks at modality in Spanish and identifies the various uses of the modals *tener que*, *deber*, *haber de* and *hay que*. She bases her semantic classes on the classification system adopted by Coates (1983). Using relative frequencies to determine which uses are most common among these modals she does not find any examples of epistemic uses of *tener que*, *haber de* or *hay que* in her data. This is surprising given the universal tendency mentioned above for these agent-

oriented modals to move towards and eventually acquire an epistemic sense. Sirbu-Dumitrescu (1988) also looks at the semantics of agent-oriented modality in Spanish but often bases the uses of these forms on the psychological motives of the speaker which, as already mentioned, cannot produce empirically reliable results.

In examining the diachronic path of Spanish obligation markers<sup>7</sup> research has largely focused on only one or two aspects of these tendencies. For example Cornillie (2007) focuses on the change from deontic to epistemic modality in two modal markers. He takes a diachronic look at *tener que* and compares the deontic and epistemic uses of this obligation marker in various contexts.

Other studies have focused on the status of these modal markers in Spanish, focusing on the aforementioned claim that modal auxiliaries are more susceptible to change than periphrastics (Coates, 1983; Westney, 1995). That being said, the Spanish auxiliary system is practically non-existent if we were to adopt the criteria put forth by linguists such as Palmer (1990) and Coates (1983).<sup>8</sup> But as Cornillie (2007: 224) states ‘unlike English modal auxiliaries, the Spanish modals have tense and mood inflections and hence, do not belong to a well-defined morphosyntactic category of modal auxiliaries.’

There have been attempts, albeit few, to identify criteria for modal auxiliaries, periphrastics and lexical verbs in Spanish. For example, Fernández (1999: 22) reviews a number of criteria that have been used to identify periphrases in general and claims the best way to identify a periphrastic construction is the combination of an impersonal

---

<sup>7</sup> The specific paths for each of these obligation markers will be discussed more in detail in section 1.4.

<sup>8</sup> Criteria for determining modal auxiliaries in English have been dubbed NICE properties which define the modal auxiliaries: negation, inversion, code, and emphatic affirmation. For a more in depth discussion on English modal auxiliaries see Coates (1983), Palmer (1987), Westney (1995), or Krug (2000).

form of the purported periphrastic alongside a phrase containing the infinitive *haber* '(to) be, have' as we see in (1.7 a,b).

- 1.7 a. {*Puede/debe/suele...*} *haber muchos problemas*  
'{There may/ must/ tends to}... be a lot of problems.'
- b. {*\*Quiere/piensa/cree...*} *haber muchos problemas*  
'{\*He wants/ thinks/believes...} to be a lot of problems'

Fernández (1999) is also aware that there are exceptions to this criterion, most notably with the construction *hay que*. Nevertheless, he considers this to be the most reliable way to recognize a periphrastic in Spanish.<sup>9</sup> Unfortunately, Fernández uses the terms auxiliary and periphrastic interchangeably so, although he gives a clear criterion for differentiating between periphrastic verbs and lexical verbs, he does not try to identify criteria for recognizing an auxiliary.

Yllera also identifies criteria for determining the status of a modal verb and lists several conditions but, like Fernández, she does not appear to differentiate between the modal periphrastics and auxiliaries (1980). These studies lead to the conclusion given by Olbertz (1998; 33) who identifies the existence of two 'true' auxiliaries in Spanish, *haber* 'have' and *ser* 'be' thereby categorizing modals like *tener que* as periphrastics. Assuming that this is true and considering Westney's statement about periphrastics, this would imply that these constructions only express epistemicity to a restricted degree. Again, this is an empirical question that requires a quantitative evaluation of the modal system in Spanish. Regardless of the answer, we now can at least identify the modal system as consisting of modal periphrastics.

---

<sup>9</sup> Fernández claims that this test does not work well with *haber que* for systematic reasons. Nevertheless a Google search of *hay que haber* turned out more than 32 million occurrences of this phrase.

## 1.2 Usage-Based Models

One way to account for the difficulty in finding clear definitions for grammatical categories such as modality (Company 2004) is through a usage-based framework. Instead of grammatical categories, usage-based models are comprised of a continuum with the prototypical uses found in the middle of this continuum and other, less common uses found towards the ends of this continuum (for more on this see Company 2004; Heine et al. 1991). Importantly, usage-based models provide ‘conceptual unification’ or a way to map language so as to account for both language function and general cognition under one model (Langacker, 2000). Langacker (2000: ix), who coined the term, also describes this model as being ‘maximalist, non-reductive and bottom up’ showing that, contrary to generative theory, it is not necessary for language structure to be economic, nor does the grammar need to be based on general rules (or categories). Rather, ‘grammar is the cognitive organization of one’s experience with language’ (Bybee, 2006: 711). Thus, this model emphasizes the relationship between experience, cognitive representation and the construction of grammar.

The focus on experience as something central to language and change is in contrast to other theories that have focused on minimalism and economy and the basic idea that expressions are generally ‘derivable by rule, [and] to list them individually would be to miss a generalization’ (Langacker, 2000: 2). Instead, a usage based model focuses on a bottom up orientation using experience as way to create connections and, ultimately, structures in the brain. Each utterance is stored not as a part of a rule or a list, but rather as an individual USAGE EVENT. A fundamental part of how grammar develops and thus changes in usage based models is largely determined by the frequency of these utterances. As Langacker (2000: 3) explains, each usage event

leaves a 'trace that facilitates [its] recurrence.' The continual recurrence of this usage event leads to ENTRENCHMENT, signifying the unit's independence as it begins to operate as an autonomous unit.

As these utterances become entrenched they are stored in close proximity to similar units. As Langacker (2000:7) points out, '[t]he occurrence of any pattern close to it in state space will tend to be facilitated' and can therefore act as an indicator of significance. Once similarities can be drawn from these structures a schema emerges and is considered a standard from which novel but similar structures will be compared (Langacker, 2000). Bybee provides evidence to this notion as she shows the development of prefabs and the high frequency of related word patterns.<sup>10</sup> She posits that the high recurrence of these prefabs in the spoken and written corpora, as well as the new constructions derived from these prefabs, are evidence for CATEGORIZATION or 'cognitive storage of particular exemplars of constructions (2006; 729).<sup>11</sup>

The formation, entrenchment and resulting categorization of word patterns and constructions through usage and frequency are important tenets of language change in a usage based model. Grammaticalization and subjectification, processes associated with diachronic language change, are intricately bound to the usage event and thus related to the entrenchment and categorization of patterns and, ultimately the formation of grammar.

---

<sup>10</sup> According to Erman and Warren (200:31), 'a prefab is a combination of at least two words favored by native speakers in preference to an alternative combination which would have been equivalent had there been no conventionalization'

<sup>11</sup> The review of usage-based models will include pertinent components only, since some aspects are beyond the scope of this study. However, for a detailed outline of the various principles of usage-based models see Biber (1999), Barlow (1999), Langacker (1999), and Barlow and Kemmer (2000).

### 1.2.1 Grammaticalization Theory

As previously mentioned, it has long been recognized that languages undergo change. Among the various types of linguistic change is grammaticalization. This type of change has been studied extensively in historical linguistics, which consequently produced a theory known as GRAMMATICALIZATION THEORY. Although the concept of grammaticalization dates back to 1912 with Meillet, who first coined the term, the theoretical framework is still relatively new. Company (2004) summarizes the various aspects and definitions the theory entails by identifying three general views on grammaticalization.

The first group, which she calls the TRADITIONALISTS, adopts a more historical approach. This approach can be summarized with a commonly used definition put forth by Kurylowicz (1976:69 [1965]), who states that '[g]rammaticalization consists in the increase of range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, e.g. from a derivative formant to an inflection one.' This definition is in line with Meillet (1912) and will be the definition adopted here. Linguists who subscribe to this view primarily see grammaticalization as a gradual process. As Heine et al. (1991) point out; however, this definition can also signify lexicalization. In an effort to confine the meaning of grammaticalization, other definitions have been posited.

Hyman (1984: 73, 83), for example, regards this process to be more narrowly defined than Kurylowicz's (1976 [1965]) definition. Hyman (1984: 73) claims that grammaticalization is 'the harnessing of pragmatics by a grammar.' Company (2004) identifies this second group as looking at this process as PRAGMATICALIZATION. Linguists in this group view grammaticalization as a change that occurs based on the contexts in

which the forms are used. That is, the more these forms are used in certain contexts the more possible it is for them to acquire new meaning (c.f. Traugott, 1982, 1989, 1999, 2003; Traugott & Dasher, 2002). Company (2004) differentiates this group from 'traditionalists' like Bybee et al. (1994), Heine et al. (1991), Bybee & Pagliuca (1985) based on where the change occurs. She states that traditionalists see this change as a gradual process that occurs to the form due to phonological erosion, loss of autonomy etc. instead of basing change on the context (Company, 2004).

The third position sees grammaticalization as a synonym of grammar (Heine et al., 1991). Linguists supporting this position claim that structure did not exist prior to grammaticalization. They also see grammaticalization as similar to any other creative cognitive process of human beings (e.g. Company, 2004). This has also been referred to by Hopper (1987) as emergent grammar.

Regardless of the position one takes, all agree that grammaticalization is a process. This process can be viewed from two different perspectives: diachronic or synchronic. Traditionally, grammaticalization has been studied from a diachronic perspective, given that it is a slow, gradual process. Adopting a diachronic approach allows for the identification of the source of the forms under examination and also illustrates the various steps these forms take as they become more grammatical (Hopper & Traugott, 2003 [1993]). From a diachronic perspective, identifying changes in progress is more difficult since the data available do not include social factors, which are often crucial to language change. Company (2004: 24) states that, in order to identify these changes diachronically, it is necessary to look at the forms within a single text to see if alternation exists. In recent years linguists have also adopted a synchronic

approach, focusing primarily on the variation within the system to determine whether a change is in progress.

Synchronic grammaticalization is often identified when two or more forms are used for what appears to be the same function. A synchronic approach allows for the incorporation of social factors given that the data is recent.<sup>12</sup> A synchronic approach also reveals the lack of discrete categories of these forms and highlights the ‘fluid patterns of language use’ (Hopper & Traugott, 2003: 3). Additionally, Labov (1972) indicates that synchronic variation is often a reflection of diachronic change so detecting changes in a modern system is essentially a window into the past. In fact, it has been pointed out in the literature that forms often retain connotations that are related to their original meaning (Bybee et al. 1994; Poplack & Tagliamonte, 2001). However, as Givón (1990 [1984]: 234) points out, a synchronic account of language may provide ‘patterns of associations and dissociation’ between the variables but a diachronic explanation is often necessary in order to explain the reasons why and how these patterns arose. Given the importance placed on both a diachronic and synchronic account of language, both approaches were adopted in this study.

An important principle that has been widely accepted as a part of grammaticalization theory is the PRINCIPLE OF UNIDIRECTIONALITY, which crucially identifies grammaticalization as a process that is unidirectional. In recent years however, there have been numerous counterexamples to this principle (Newmeyer, 1998; Campbell, 2001), which have consequently called the theory of grammaticalization into question (Fischer, Norde & Perridon, 2004).

---

<sup>12</sup> A more detailed look at the benefits of diachronic study vs. synchronic study will be offered in Chapter 2

Haspelmath (2004: 3) confronts this issue and admits that there are a limited number of counterexamples, like the change from an affix to a clitic in Continental Scandinavian, but that, overall, grammaticalization is unidirectional. In addition, he adds that the goal in historical linguistics is to ‘identify universals of language change’ and classifies unidirectionality in grammaticalization as one of the most important universals in language change given that it is so widespread (2004: 35). Haspelmath (2004) also discredits many of the criticisms of claims of unidirectionality citing several reasons as to why they were not counterexamples. Other studies have addressed the principle of unidirectionality and its validity (c.f. Taeymans, 2004, Hopper & Traugott, 2003 [1993]) where general results may vary slightly from Haspelmath (2004) but ultimately all arrive at the same conclusion: the general pattern of grammaticalization is unidirectional.<sup>13</sup>

Given the different interpretations as to what grammaticalization theory entails and the counterexamples to the principle of unidirectionality, there has been debate as to whether or not grammaticalization is a theory (Fischer, Norde, & Perridon, 2004). Haspelmath (2004) argues that, although grammaticalization theory encompasses many approaches, it can still be a theory. He equates it to ‘evolutionary theory’ stating ‘evolutionary theory...is not one single monolithic system either, but describes a range of related approaches and basic issues in the area of historical biology’ (2004: 23). Company (2004: 1) also sees grammaticalization as a theory and argues that if one adopts the functionalist perspective of what grammaticalization is—a ‘predictive line of investigation that accounts for many types of linguistic change’—then one can say it is a theory. For the sake of the present paper, grammaticalization will be considered a

---

<sup>13</sup> For more on the discussion of unidirectionality and these counterexamples see also Hopper and Traugott (1993) Tabor and Traugott (1998), Traugott (1995) and Narrog (2005).

theoretical framework and furthermore, the definition put forth by Kurylowicz (1975:52 [1965]) will be the definition adopted.

Research within the grammaticalization framework has concluded that areas of grammar that refer to specific conceptual domains are more apt to undergo grammaticalization. Bybee (1998: 258) asserts that only a small set of words in a language are susceptible to grammaticalization and ascertains that they typically relate to notions about human experience across cultures: she offers as specific examples of words referring to spatial concepts, such as the body, movement in space etc. Traugott & Dasher (2002: 3) similarly identify areas as primary hot spots for grammaticalization but add to the list by including aspect (*have, finish*) and modality (*want, will*). The changes that typically occur to the last group –modality- are further described as a general semantic-pragmatic shift. This occurs as they acquire a more epistemic and thus subjective meaning, a process known as SUBJECTIFICATION (Traugott, 1989; Bybee et al., 1994; Tagliamonte, 2004; Narrog, 2005; Tagliamonte & Smith, 2006; Tagliamonte & D'Arcy, 2007).

An important aspect in grammaticalization theory to consider is frequency. As mentioned above particular consideration is given to frequency in a usage based model as it is through these high frequency forms that we find language change like grammaticalization. These forms become more grammatical through processes such as decategorialization, phonetic reduction, reanalysis, semantic bleaching or emancipation from the source form. It is these mechanisms that, when linked together, comprise the cline of grammaticalization (Hopper & Traugott, 2003 [1993]). However, not all these processes are required in order for a construction to be considered as undergoing

grammaticalization. Considering that grammaticalization is a gradual process, it is difficult to witness a form that has gone through the entire process.<sup>14</sup> Nonetheless, there are certain examples, such as the Latin verb *habere* 'have' or the English construction *going to* that have undergone many of these stages, which will be discussed in greater detail in section 1.3.1.1.

### 1.2.2 Mechanisms in Grammaticalization

If we look at the example *going to* we see that, in its original sense (1.8), this construction had a sense of direction or movement (Hopper & Traugott, 1993; Bybee, 2003). At the same time it is also evident that direction/movement is not the only connotation of this construction. Intention can also be inferred in this context in that the listener assumes that the speaker intends to travel to London in the future to get married. It is the ambiguity of this sentence that can lead to the grammaticalization of the construction.

1.8 I am **going to** London to marry Bill. (Hopper & Traugott, 1993: 2001)

The fact that intention can be inferred is not enough, however, because in order for

inferences to play a significant role in grammaticalization, they must be frequently occurring, since only standard inferences can plausibly be assumed to have a lasting impact on the meaning of an expression or to function cross-linguistically (Hopper & Traugott, 2003; 82).

We know in the case of *going to* inference did play a significant role, since we now have

1.9 We are **going to** get married in June (Bybee, 2003; 147)

1.10 These trees are **going to** lose their leaves. (Bybee, 2003; 147)

---

<sup>14</sup> Again, certain features such as affix binding, clause combining are not discussed as they are outside the bounds of the current paper.

examples such as 1.9 and 1.10. Of importance is the fact that in 1.9 we can still extract a sense of movement or direction, whereas in 1.10 this sense has been completely lost.

This loss, known as SEMANTIC BLEACHING, or sometimes DESEMANTICATION, is also the result of increased frequency and entails the general shift of a construction into new contexts. Semantic bleaching can sometimes indicate the complete loss of semantic content such as in the case *haber* in Spanish.

During the 13<sup>th</sup> and 14<sup>th</sup> centuries, this form was still being used to express possession as we see in 1.11, although it appeared more frequently as an auxiliary and as an expression of obligation.

- 1.11 *O Antigono, tu as muchos amigos, & de çaga tu as muchos grandes hombres* (Fernández de Heredia, Juan, 1379 – 1384: Corde)  
'Oh Antigono, you have many friends, and behind you, you have many great men.'

By the 15<sup>th</sup> century, however, this meaning cannot be found in the data.<sup>15</sup>

Semantic bleaching does not always entail loss, however. Crucially, the ambiguity seen in *going to* in 1.8 and the resulting POLYSEMY of this construction (movement into new contexts) leads to autonomization, even though they are semantically and historically related. In this way, the two semantic interpretations can exist for long periods of time without falling into disuse, like *going to*.

Although polysemy indicates that the construction has emerged in new contexts, this development does not occur in order to fill a 'functional gap' (Hopper & Traugott, 2003). Thus the semantic shift commonly entails moving into the territory of another

---

<sup>15</sup> Nevertheless there are still remnants of the possession meaning in its auxiliary form, particularly in sentences like:

porque muchos temores adelantados han abierto las puertas a la ofensa /because many anticipated fears have opened the door to insult. (CORDE, Márquez, Fray Juan, 1612 - a 1625)

form. This process commonly known as RENEWAL, or what Tagliamonte (2007) refers to as longitudinal renewal, is an alternate way of saying the same thing and is ubiquitous in language (Hopper & Traugott, 1993). Take for example the Spanish construction *ir a* ‘go(ing) to’ which has followed a grammaticalization path similar to that of English *going to* (Bybee, Pagliuca, & Perkins, 1991; Bybee, Perkins, & Pagliuca, 1994).<sup>16</sup> Once a meaning of futurity is established the construction starts to appear in contexts designating a future action. This causes the construction to go into direct competition with the Spanish synthetic future tense like we see in 1.12 and 1.13. This competition, known as layering, is often an indication of grammaticalization.

1.12 *Sancho, cuanto quisieres, que yo no te iré a la mano, pero mira lo que vas a decir.* (CORDE, Cervantes Saavedra, Miguel de, 1615)  
‘Sancho, however much you should want, I will not go with you and hold your hand, but look this is what you’re going to say’

1.13 *Buelue, Zara, por tu honor, mas diras que es ignorancia,* (CORDE, Anónimo, 1605)  
‘Return, Sara, because of your honor, but you will say it is because of ignorance.’

This variation can also indicate the gradual replacement of one form by another.

Layering has been evidenced in a wide variety of research as a possible indication of a change in progress (e.g. Tagliamonte & D’Arcy, 2007; Tagliamonte, 2004; Silva-Corvalán, 1985).

Development along the grammaticalization cline can also involve phonological REDUCTION. In the case of *going to*, we see that in 1.14 the construction has phonologically reduced to *gonna*. It is also through this phonological reduction that the

---

<sup>16</sup> This specifically refers to what has been discussed up to this point i.e. pragmatic inference whereby a movement/direction meaning grammaticalizes and acquires a future and intention meaning. The analogous paths of change with these two forms does differ in many ways like, for example, with reduction.

evidence of semantic DIVERGENCE (from the original sense of movement/direction) is most conspicuous. As Bybee (2003) argues, the original form is not the structure that has undergone phonologically reduction. Rather, it is the form that denotes intention and future that is the grammaticalized construction. This is evident because sentences that only imply direction cannot be used with *gonna* as in 1.15.

1.14 I'm **gonna** help you later.

1.15 I'm **going to**/\***gonna** the store.

Phonological reduction as an effect of frequency has also been largely attested in the literature. For more examples of this see Bybee (2001), Company (2006), and Tagliamonte (2004).

Empirical support also identifies DECATEGORYLIZATION as another common characteristic of grammaticalization (Company, 2006; Torres, 2006; Cornillie, 2007). Torres Cacoullos (2006) illustrates this process in her diachronic study of the Spanish collocation *a pesar de* 'despite' showing how *pesar*, originally a noun in Old Spanish, has grammaticalized and diverged, and the grammaticalizing form has lost its nominal trappings as an effect of frequency. This loss of certain attributes –syntactic or morphological- results in the inability to 'identify it as a full member of a major grammatical category such as noun or verb' (Hopper & Traugott, 2003; 107). In the case of *a pesar de* the decategorialization of *pesar* as a noun can be credited to the high frequency of the construction *a pesar de* relative to the frequency of the lexical noun *pesar* and the subsequent loss of nominal trappings such as plural marking, and determiner and adjective modification (Torres, 2006; 38). The fusion of these previously independent morphemes is the result of the re-categorization of the construction as a

single unit phonologically, semantically and syntactically (Bybee, 2006). As we will see below, decategorialization is particularly important as it is one of the processes where grammaticalization and subjectification overlap.

What emerges from this discussion is that language use and thus frequency of certain structures are the locus of processes like grammaticalization. In the section that follows it will be seen that frequency is also a crucial factor in subjectification.

Additionally, many of the same factors revealed to be central to grammaticalization are the same that produce the ideal environment for subjectification to ensue.

### 1.2.3 Subjectification

There are two different accounts of what subjectification entails. Those that follow Langacker (1999a, 297) define subjectification as a 'shift from a relatively objective construal of some entity to a more subjective one.' He explains an entity is construed subjectively 'to the extent that it functions as the subject of conception without itself being conceived' (1999a: 297). This definition in fact works well with examples like *tener que* in 1.16 and *dale* in 1.17 where the speaker is essentially offstage and the statement is indeed subjective.

1.16 *El enemigo **tiene que** estar precisamente dentro de este círculo* (Torrente Ballester, Gonzalo, 1972: Corde)  
'The enemy must be precisely among this group. (Based on given evidence)'

1.17 ***-Dale** que dale, pero qué pesada eres.* (Company, 2006)  
'Dale over and over again. What a crashing bore you are.'

However, as Narrog (2005a) points out this does not always hold true since an example like *I'm sure* can be subjective and the speaker is essentially onstage and conceived of, which is in direct contradiction to Langacker's (1999) definition of subjectification. For

the present paper a perspective that disallows the speaker to be onstage could possibly exclude valid instances of subjectification.

A more general account of subjectification is adopted by those who follow Traugott's (1989) account where she identifies three tendencies in semantic-pragmatic change, all of which represent different levels of subjectification. She describes the first tendency as a shift from meanings centered on the 'external described event' to a meaning that is more focused on the 'internal described situation' (1989: 34). She explains that the internal situation often entails evaluations and perceptions. As a construction continues to undergo subjectification, another change of meaning can be observed where there is a shift from the 'external or internal situation to those meanings based in the textual and metalinguistic situation' (1989: 35). She defines the textual or metalinguistic situation as a shift of the construction towards contexts where it has more of a speech-act like function. She develops this idea by incorporating the third tendency where 'meanings tend to become increasingly based in the speaker's subjective state/attitude toward the proposition' (1989: 35). Although this conceptualization of the process is, on the surface, similar to Langacker's definition and could indicate only a 'terminological' difference as Langacker (1999b) claims, Traugott (1989: 188) disagrees pointing out that, for her, subjectification can only be a diachronic phenomenon where as Langacker's perspective of subjectification is synchronic.

Cornillie (2007) finds these two views complementary illustrating that in such cases as the subjectification of Spanish *prometer* 'to promise' and *amenazar* 'to threaten' both views are reflected: the synchronic attenuation of these forms supports Langacker's (1999b) view; and three diachronic stages supported by Traugott's (1989)

perspective can also be identified. Cornillie's interpretation of subjectification is likely the most encompassing, and as a result will be the view adopted in the present paper.

By adopting both a diachronic and synchronic perspective on subjectification we can now attempt to identify the various points in subjectification as the forms develop. Several clearly identified processes have been acknowledged in subjectification, namely increased transparency, which involves decategorialization (Langacker, 1999; Torres, 2006; Company, 2006, Cornillie, 2007), the tendency to appear more with first and second person (Aaron & Torres, 2005 and references therein) or, in the case of epistemic modality, the tendency to occur with third person animate subjects (Pietrandrea 2005; Bybee and Fleischman, 1985; Coates, 1983), and syntactical flexibility (Torres & Schwenter, 2004; Company, 2006). It has been particularly challenging for linguists to operationalize the various degrees of subjectification (Traugott 1999, 1995; Torres & Schwenter 2004; and Company, 2004; Aaron & Torres, 2005), although there have been several attempts, most notably Company (2004), Torres & Schwenter (2004) and Aaron & Torres (2005).

Torres and Schwenter (2004) outline a specific way to measure subjectification using the Spanish construction *a pesar de* 'despite' as the subjectifying form. They identify several measures as ways to determine the evolution of subjectification, although their analysis is confined to connectives. Thus, although these measures are indeed quantifiable, they do not provide a general account of how to measure subjectification in language, but rather a very specific account of how to measure subjectification in connectives.

Company (2004) correlates subjectification with syntax in her study on discourse markers in Spanish such as *dizque*, *tate* and *dale*. She demonstrates how these forms increasingly become constructions expressing the speaker's point of view, while at the same time, displaying syntactic isolation and cancellation of syntax. She clarifies that syntactic isolation happens when an originally rigid form becomes less rigid and is able to appear separate from the rest of the sentence, usually indicated by commas, as seen with *ándale* in 1.18. She also emphasizes that a form in syntactic isolation can be a 'whole predication in and of itself' which she would define as being 'stripped of syntax' or the cancellation of syntax (2004: 14).

- 1.18 *Tú cállate. Tráenos algo, **ándale*** (Carlos Fuentes, 1962 in Company, 2004)  
'You shut up. Bring us something. Andale.'

She effectively operationalizes this process through the identification of various stages of subjectification as these verbs make their way towards a state of high subjectivity creating the shift into discourse markers. However, like Torres and Schwenter (2004), this operationalization is specific to discourse markers, which makes it difficult to apply to other domains. What both Torres and Schwenter (2004) and Company (2004) highlight, nevertheless, is that there is a general dearth of research quantifying subjectification.

Besides these specialized perspectives of subjectification related to discourse markers and connectors, there is also an example in the literature of the subjectification of grammaticalizing forms. Aaron and Torres (2005) measured the subjectification of *salirse* 'to leave' by employing a variationist method. They hypothesize that if *salirse* is indeed subjectifying, then it will likely occur more often with factors that reflected that

process. To operationalize the process they chose three factors that might reflect subjectification: co-occurrence of dative pronouns, grammatical person, and relationship to the speaker (in terms of distance). Through the utilization of multivariate analyses, they were able to measure the effect of these factors and identify patterns of subjectivity in *salirse*.

Results indicated that this form is more likely to co-occur with first person, as well as with referents close to the subject. This study appears to be the first and only to apply a variationist method to measure subjectification. At the same time, it does not negate that clear measures of subjectification are still lacking in the field. Consequently, the present analysis will be based less on the degree or phase of subjectification of these forms, but rather highlight some of the clearly identified characteristics of the process and correlate them with grammaticalization as well as with the constructions *haber de*, *tener de*, *hay que* and *tener que*.

#### **1.2.4 Subjectification and Grammaticalization.**

In the literature there has been general acceptance that, in many cases, subjectification is part of grammaticalization (Bybee et al., 1991; Bybee et al. 1994; Hopper & Traugott, 2003, Brinton & Traugott, 2005; Narrog, 2005a). This is especially accounted for in the postulation of universal paths of grammaticalization, particularly that of agent-oriented modality (e.g. obligation) to epistemic, or speaker oriented, modality (Bybee et al. 1994, Narrog, 2005a).

Since grammaticalization involves shifts toward more abstract, less referential, markers, the prime function of which is to represent the speaker's perspective on the situation or to get others to do things it is necessarily the case that subjectification is characteristic of grammaticalization (Brinton & Traugott, 2005; 108).

As noted by some of the studies reviewed above, there are certain domains where these two processes overlap, primarily with decategorialization, pragmatic inferencing and the general movement towards a more abstract meaning. This, like in grammaticalization, is evidenced with the rise in frequency of a construction. Take for example the Spanish construction *a pesar de* 'despite' and its resultant shift from a noun *pesar* to a connective (Torres, 2006). With the same example, Torres and Schwenter (2004) use this shift from noun to connective as an indication of subjectification. Similar examples can be found in Aaron and Torres (2005), Narrog (2005a), Cornillie (2007) and Tagliamonte and D'Arcy (2007).

Nevertheless, as Traugott and Dasher (2002: 89-90) point out, 'subjectification is typical of semantic change in general and is not limited to grammaticalization'. Essentially, subjectivity is ubiquitous in language and, as Finegan (1995: 6) states, there is an 'emerging view of discourse as an instrument not solely [...] designed for communicating ready-made content, but as an expression of self and, in part, its creation. Therefore, it is difficult to confine subjectification to a sub-process of grammaticalization when in fact it is present in many parts of language since expression of self is the ultimate goal (Finegan, 1995).

This can be seen in the subjectification of the Spanish form *claro* as it moves from an adjective to a discourse marker. Ocampo (2006) is quick to recognize similarities between grammaticalization and subjectification but also emphasizes some of the contradictory characteristics, namely the general path of grammaticalization from lexical to grammatical, which he specifies would indicate a movement towards syntax. Whereas he points out that in subjectification, at least in the subjectification of discourse

markers, the path of these constructions is out of syntax and towards discourse (2006: 317). His study illustrates that subjectification can occur outside of grammaticalization and also raises some valid questions in drawing the line between the subjectification of discourse markers and grammaticalization. Nevertheless, these questions are beyond the scope of this paper since *tener de*, *tener que*, *haber que* and *haber de* have not become discourse markers.

In the present paper, as will be seen below, any subjectification that occurs is ultimately a small part of a bigger process: grammaticalization. We assume this not only due to the various examples of these constructions show of having gone through certain stages of grammaticalization, which will be discussed in more detail in section 1.4, but also because of the consensus in the literature that subjectification within agent oriented modality, specifically with periphrastics, is related to grammaticalization (Traugott, 1989; Bybee et al. 1991; Bybee et al. 1994; Narrog, 2005).

### **1.3 The Obligation System in Spanish**

As previously mentioned, there have been certain universal paths of grammaticalization postulated by linguists such as Bybee et al. (1994), Bybee et al. (1991) and Hopper and Traugott (1993). These linguists posit that constructions denoting a notion of possession develop into obligation markers and eventually acquire an epistemic or subjective meaning. The four constructions *tener que*, *haber que*, *haber de* and *tener de* all derive from such a source in that *tener* 'have' and *haber* 'have' historically conveyed a notion of possession (for more on the historical development of these forms see Yllera, 1980). A semantic pragmatic shift in each of these collocations could imply grammaticalization and even subjectification if the polysemy includes epistemic modality. Furthermore, an apparent emancipation of these constructions from

their source form could also be indicative of grammaticalization, while their decategorialization could designate both grammaticalization and subjectification. Finally, the fusion of independent morphemes to form collocations is also a possible indication of grammaticalization.

Little work has been done on the grammaticalization of these expressions of obligation in Spanish. Taking each of these apparent changes into account, we will examine these constructions and determine whether these forms are undergoing grammaticalization. In addition, we will look at the constructions and the ostensive shift towards a more subjective reading as possible corroboration that these forms are likewise undergoing subjectification.

### 1.3.1 The (Non)grammaticalization of These Constructions

One of the first measures of grammaticalization that we see in these constructions necessarily occurs prior to them becoming expressions of obligation. In order for each construction to grammaticalize from a construction expressing possession, semantic bleaching needs to occur where an obligation reading can be inferred from the context (Hopper and Traugott, 1993). We can see in examples 1.19-1.21 that both a possession and an obligation reading can be inferred. Although in present day Spanish the verb

- 1.19 *y yo soy el que **tengo** necesidad **de** ser regido* (Sahagún, 1576-1577: Corde)  
'and it is I who has the need to be governed.'
- 1.20 *No **tengo** más **que** dezir.* (Sahagún, 1576-1577: Corde)  
'I don't have more to say.'
- 1.21 *que no **han** menester más **de** ver un no sé qué y sin más mirar ni enterarse de cierto en lo que era y cómo era* (Cárdenas, 1591: Corde)  
'that they don't have more need than to see a I don't know what and with more than seeing or finding out for sure about what it was and how it was.'

*haber* has lost its possession meaning, during the earlier centuries it was still used as a possession marker. Linguists (e.g. Cornillie, de Mulder, Van Hecke and Dieter Vermandere, 2009; Traugott & Dasher, 2002) suggest that when the object is something that is not yet possessed semantic bleaching will occur.

In the case of *tener que*, there is disagreement as to whether the development of this obligation marker was the result of the aforementioned process of semantic bleaching (e.g. Cornillie et al. 2009, Yllera, 1980) or whether it came about from the semi-transitive examples of *tener de* (Cornillie, 2007: 228) in which the construction occurs in contexts with both the preposition *de* and the conjunction *que* as is seen in the example 1.22 given by Cornillie et al.<sup>17</sup>

- 1.22 *si el procurador fuere rebelde que non sea restituído el señor del pleito maguer que diga que el procurador non **tiene de que** pagar.* (Alfonso X. *Siete partidas*. 13<sup>th</sup> century)  
'if the attorney opposes the fact that the lawyer be refunded although he says that the attorney does not have of to pay. (i.e. something for which to pay)'

As these constructions start to occur in contexts where obligation may be inferred we also see each marker undergo decategorialization and a subsequent shift into agentive modality. We can clearly see the difference in examples 1.23-1.24 in comparison to 1.25-1.26 where both *tener* and *haber*, transitive verbs, have evolved into semi-auxiliaries and lost some of the properties that would otherwise categorize them as full verbs. For example, when these verbs occur in this construction they can no longer take a nominal complement. Additionally, the subject of the construction seems

---

<sup>17</sup> Although Cornillie (2007) suggests this as a possible path of development of the obligation marker *tener que*, in Cornillie et al. (2009) he contradicts this claim by assuming that *tener que* developed through the semantic bleaching of contexts where it expressed possession.

bound to the infinitive. That is, the subject of the two is always the same.<sup>18</sup> Both of these seem to point to decategorialization. This decategorialization, as was mentioned above, is not only a defining characteristic in grammaticalization but can also denote subjectification. As indicated in Torres (2006), decategorialization can often lead to autonomization. As the constructions lose their status as full verbs, we also see that the individual morphemes become more and more bound, indicating autonomization, and a possible divergence from the source form.

- 1.23 *O Antígono, tu **as** muchos amigos, & de çaga tu **as** muchos grandes hombres* (Fernández de Heredia, Juan, 1379 - 1384)  
 ‘Oh Antígono, you have many friends, and in front of you, you have many great men’
- 1.24 *Bien, bien (dije yo entonces) noticia **tengo** de estas fiestas;* (CORDE, Gómez de Tejada, Cosme, 1636)  
 ‘Well, well (I then said) I have news of these celebrations.’
- 1.25 *Señor, ¿con qué mandamiento te **he de** matar?* (CORDE, Quirós, Francisco Bernardo de, 1656)  
 ‘Sir, with what order do I have to kill you?’
- 1.26 *Yo **tengo de** ir con vos.* (CORDE, Calderón de la Barca, Pedro, 1629)  
 ‘I have to go with you.’

The first piece of evidence is the gradual disappearance of the source form *haber* as a notion of possession. This notion eventually becomes obsolete while the grammaticalizing construction *haber de* continues to be used. Although *tener* has not become obsolete, we do see a continuous rise in frequency of the collocation *tener que* (see section 2.1.1 for frequency counts of this collocation).

Fusion of the morphemes and the infinitive is also noticeable with all the constructions as they begin to show a lack of syntactical flexibility in terms of what can

---

<sup>18</sup> This could be the loss of agent control Company (2005) identifies as one of the stages in the subjectification of discourse markers.

occur between the construction and the infinitive. As we can see in 1.27-1.28, there was some flexibility early on, where clitics could be sandwiched between the collocation and the infinitive although even in the 17<sup>th</sup> century these examples were already rare in the corpus. This is a strong indication that early on these units had already become obligatorily attached to an infinitive.

- 1.27 **e de le** *responder con toda mayoria e superioridad* (CORDE, Fernández de Córdoba; 1625)  
'I have to respond with superiority and the majority'
- 1.28 *Siendo esa vuestra opinión, ya no **tengo que os**\_decir.* (Calderón de la Barca, Pedro, 1649)  
'This being your (pl) opinion, I do not have to tell you (pl) anything.'

This layering of forms is another measure of grammaticalization and, as previously mentioned, frequently indicates the possible replacement of one form for another. This happens with *tener de* as its use steadily declines throughout the 18<sup>th</sup> century until disappearing completely in the early 19<sup>th</sup> century. Interestingly, *haber de* also appears to have declined albeit in more recent times, according to the literature. Cornillie (2007) points out that *haber de* is used more often used in writing than in speech.

- 1.29 *de la doctrina cristiana que todos los niños **han de saber**...* (Anónimo, 1626, CORDE)  
'from the christian doctrine that all children have to know...'
- 1.30 *yo amé ya no **tengo que** dudar, ni tú **tienes que** saber...* (Calderón de la Barca, Pedro, 1635, CORDE)  
'that I loved already, I don't have to doubt, and you don't have to know.'
- 1.31 *porque a riesgo de mi vida, **tengo de** saber quién sois.* (Calderón de la Barca, Pedro, 1635, CORDE)  
'because at the risk of my life, I have to know who you are.'

Without a comprehensive study, however, it is hard to confirm or refute his claim. At this point we can posit that these forms are grammaticalizing. Now the question is whether there is also an indication of subjectification.

### **1.3.2 The (Non)subjectification of These Constructions**

Previous research on subjectification identifies several factors that may indicate the subjectification of these obligation markers. Based on the description of the decategorialization of each of these forms in the previous section, it is postulated that these constructions have entered into the subjectification process. This would indicate a semantic-pragmatic shift towards epistemicity. In the case of deontic modality, the agent who expresses the obligation is usually not involved in the proposition (Bybee et al., 1994). Rather, the domain of obligation is based on social norms or physical necessity that is external to the speaker (Bybee et al. 1994; 201). As these forms strengthen pragmatically and the domains shift, they acquire an epistemic use whereby the attitude of the speaker towards the proposition is now central to the utterance. This would be a prime example of subjectification in that it shows the semantic reallocation of a construction that is based on things external to the speaker and its eventual evolution towards the expression of the value of truth the speaker has towards the proposition (i.e. epistemicity). However, in order for a transformation like this to arise, remember that high frequency of occurrence is indispensable.

If we look at examples of *haber de* and *tener de*, the older and more common constructions in the earlier centuries, from the 17<sup>th</sup> century we see that *haber de* has already acquired an epistemic meaning. In fact, this form has an epistemic notion as early as the 14<sup>th</sup> century (1.32) which is a strong indication that this verb came into Spanish (from Latin) already having gone through subjectification. In fact, as Cornillie

(2007: 226) points out, due to the future reading of *habere* in Latin (*cf. cantare habeo > cantaré*) it is possible that a future connotation precedes the deontic reading.<sup>19</sup>

Besides *haber de*, we also see epistemicity expressed with *tener de* (1.33) early in the 17<sup>th</sup> century, although Cornillie (2007: 229) finds an example as early as the 16<sup>th</sup> century. The fact that these constructions have decategorized (e.g. no longer can take nominal complements) and both forms are comprised of a subjective function point to the possibility that they are undergoing subjectification. Additionally, the continued layering of these forms could be an indication that they are still undergoing grammaticalization and subjectification. This study aims to determine whether these constructions continue to grammaticalize and whether or not they have entered into subjectification.

- 1.32 *¿Sabéis cuál dellos **ha de** ser muerto?* (CORDE, Anónimo, c 1400 - 1498)  
'Do you (pl) know which one of them must be dead?'
- 1.33 *Mi esclava **tiene de** ser, que es ley de Dios inuiolada, nunca vista dispensada ni con hombre ni muger.* (CORDE, Valdivielso, José de 1622)  
'She must be my slave, which is the unviolated law of God, never seen dispensed nor with a man or a women.'

In addition to the aforementioned characteristics of subjectification, recall that Traugott (1989) identifies three tendencies she found to be common in constructions undergoing subjectification. In the same study she correlates these tendencies with the development of deontic (obligation) and epistemic modality. She proposes a path of subjectification specific to obligation (table 1.1). Based on her postulation, each of the constructions in Spanish have already seen a shift from an 'external described event', which in this case is possession, to an 'internal described situation,' which could be

---

<sup>19</sup> This infers that, in order to get a clear representation of the subjectification path of *haber de*, it may be necessary to consider data from Latin.

attributed to the shift from possessing a tangible object to possessing something more intangible, like a secret (Traugott, 1989:34). Based on table 1.1, we can see that each of these obligation markers is likely reflecting uses that represent the second or third tendency.

Recall that the second tendency is identified as a shift from meanings centered on the external or internal situation to a meaning centered on the textual or metalinguistic situation (1989: 35). Traugott (1989:37) claims that, in regards to these modals, tendency II is defined by the shift of possession to obligation. This presumes then that these constructions have already undergone the shift from tendency I to tendency II and

Table 1-1. Traugott's (1989) path of subjectification of modals (re-printed with permission)

Tendency		I	II	III	III
stage	main verb	pre-modal	deontic	weak epistemicity habitual prophetic/rel. future	strong epistemicity

all should have meanings associated with tendency III, which would imply that each construction is being used more often as a way to express the speaker's subjective attitude towards the proposition (epistemicity). According to her path of development for these modals, each construction is somewhere between weak and strong epistemicity (as can be seen in table 1.1). The question is, then, where in the subjectification process is each of these markers? One of the ways to determine this is to take a variationist approach to language change, which is discussed in more detail in the following section.

## 1.4 A Variationist Perspective

One way that linguists have been able to examine grammaticalization (and subjectification) within and across languages is through a variationist approach. A variationist perspective looks at variation in language as structured and heterogeneous. The idea that language is non-uniform was first put forth by Weinreich, Labov & Herzog (1968) where they state that:

a model of language which accommodates the facts of variable usage and its social and stylistic determinants not only leads to more adequate descriptions of linguistic competence, but also naturally yields a theory of language change that bypasses the fruitless paradoxes with which historical linguistics has been struggling for over half a century (1968; 99).

Importantly, this structured heterogeneity indicates that the variant chosen by the speaker is not a random choice but, in fact, conditioned by certain social and linguistic factors. Historically, this type of variation has been largely ignored since other branches of linguistics primarily focus on those variables that are predictable by rule (Labov, 1972). This, in essence, disqualifies a variant that is not used categorically in one context or another. Through the adoption of this approach and the use of a standard statistical model that explains this variation, linguists have created a way to study these variable forms.

### 1.4.1 The Variationist Method

Working within variation theory, Labov (1972, 1994, 2004) has established certain compulsory components to examining variation in the system. One of these is the identification of the linguistic or dependent variable. These are the forms speakers use that appear to be in free variation or competition (layering) as noted above. That is the 'class of variants among which speakers alternate in the expression of a given meaning or function in discourse' (Poplack & Tagliamonte, 1998: 88). For example, in the present

study the linguistic variable is obligation, which includes four variants: *tener que*, *haber de*, *hay que*, and *tener de*. These variables appear to have the same functions given the examples below (1.34-1.37). However, from a variationist perspective we can see that this is not true.

- 1.34 *Préndanme; córtanme la cabeza, que yo cumplo con mi oficio, y he de decir\_la verdad.* (Barrionuevo, Jerónimo de, 1654-1658, Corde)  
'Catch me, cut off my head but I do my job and I have to tell the truth'
- 1.35 *Hermano, yo os tengo de decir\_la verdad; no sé qué dicen de un indiano.* (Vega Carpio, Lope de, 1632: Corde)  
'Brother, I have to tell you all the truth, I don't know what they say of a returning Spaniard who made his fortune in America.'
- 1.36 *...de esto no ay que decir\_aqui nada, porque tiene su lugar propio* (Jiménez Patón, Bartolomé, c 1604 – 1614: Corde)  
'...of this (you) must not say anything else, because it has its own place.'
- 1.37 *Con esto no tengo que decir más; harto siento lo dicho, y más cansar a vuesa merced [usted].* (Góngora y Argote, Luis de, 1613 –1626: Corde)  
'Of this I don't have to anything else, I am fed up with it...'

Bayley (2002) also identifies the importance of the PRINCIPLE OF QUANTITATIVE MODELING which essentially allows for the close examination of the 'forms that a linguistic variable takes, and [the notation of] what features of the context co-occur with these forms' (2002: 118). Such analysis can be done through a statistical program and probabilistic data. As Bayley explains, these probabilities measure the effect each of the independent variables has on the linguistic variable. For the current study, this will be done through the use of Goldvarb (Cedergren & Sankoff, 1974, Sankoff, 1988), a statistical multivariate analysis program that takes the idea of structured heterogeneity (Weinreich et al., 1968) and quantifies it (Tagliamonte, 2006).

These types of statistical programs allow for probabilistic predictions of the conditioning effects of the independent variables on each of the obligation markers. That is, we can examine from a quantitative perspective the surrounding linguistic context to identify which variables (linguistic or extralinguistic) co-occur with each obligation construction (Bayley, 2002). In addition, Goldvarb provides the relative frequencies of occurrence of each of these markers in a given context, which also provides a general account of the status of these constructions.

In addition to defining the linguistic variable and choosing independent variables it is also important to exclude any context that is categorical or near categorical with one of the variants. Labov (2004) describes these contexts as neutralizations. Excluding categorical categories allows for a clear analysis of the dependent variable and not the various other subclasses that these forms may comprise (Labov, 1972: 72).

#### **1.4.2 Comparative Method**

The adoption of the comparative method into variationist linguistics has provided sociolinguistics with a valuable tool for reconstructing languages, determining source languages in contact situations, comparing cross-dialectal varieties or two different time periods to determine the status of grammaticalizing forms, as well provide a way to look at variation from a more fine-grained perspective. Although historically the comparative method was criticized for its assumption of uniformity in the source language, its adoption into variation theory has largely resolved this limitation. In addition, this method has led to a shift from focusing on surface forms and frequency distributions as ways to identify linguistic change to primarily focusing on the variable grammar and probability values. Through the comparison of these probability values across time periods, possible sources of change in the grammar can be identified. The identification of these

sources can elucidate possible directions of change at a micro-level and corroborate assertions on the macro-level thus substantiating universal claims of paths of development as well as providing new information on how these changes take effect. This is particularly pertinent to Spanish modality given that very little quantitative research has been conducted on this system and despite the recognition that modality systems in other languages are in flux.

### **1.5 The Present Study**

Given the large amount of research that has been conducted on modality over the last several decades, it is surprising that only a small amount of that research has been conducted on the Spanish modal system. Of the literature that does focus on modality in Spanish, very little of it recognizes semantic differences between the aforementioned constructions (c.f. Rabadan, 2006). Furthermore, to my knowledge, there have not been any studies in Spanish that examine modality and its diachronic development using a quantitative variationist perspective. One may assume, however, that this can be accounted for by the research conducted on other languages given the recognition by linguists that modality tends to follow a universal path of change (Lyons, 1977; Coates, 1983; Bybee et al. 1994; Bybee & Fleischman, 1995; Palmer 2001). Nevertheless, this evidence is widely based on studies that have (i) not taken a variationist sociolinguistic perspective (e.g. Lyons, 1977, Bybee et al., 1994), (ii) limited analyses to the comparison of relative frequencies (e.g. Myhill, 1995a; Cornillie, 2007) or (iii) did not use oral speech to consider possible sociolinguistic factors that have contributed to these changes (e.g. Bybee et al., 1994).

As a result of the narrow scope that has been utilized thus far, little evidence has surfaced on how these obligation markers are used in both text and oral speech. This

study seeks to fill this gap of knowledge by examining at the development of modality in Spanish both diachronically and synchronically through a comparative sociolinguistic approach (Poplack & Meechan, 1998; Poplack & Tagliamonte, 2001; Tagliamonte, 2002; Tagliamonte & Smith, 2006). The aim is to uncover the linguistic and extralinguistic factors that condition the use of one expression of obligation over another while at the same time substantiating universal claims of development in modality. In addition to answering this overarching question, a review of the questions central to this study are listed below:

- What are the semantic differences among these four obligation markers?
- Have these forms continued to grammaticalize?
- If so, what is the path of grammaticalization of each marker?
- Have they also started to undergo subjectification?

This dissertation will be organized as follows. Chapter 2 explains the methodologies used, while the third chapter reviews the results found in the multivariate analyses, and comparative analyses. Chapter 4 is dedicated to analyzing the results of the multivariate analyses, the comparative analyses and the relative frequencies of these constructions in each century. The final chapter presents the conclusions and implications of the study, as well as suggested directions for future work.

## CHAPTER 2 METHODOLOGY

### **2.1 Corpora**

The region chosen for this dissertation was Mexico, since I have strong connections to a community in Guanajuato, Mexico, making it reasonable to conduct the sociolinguistic interviews there. The decision to limit my research to Mexico was primarily based on previous research on modality (Tagliamonte, 2002; Tagliamonte & Smith, 2006; Tagliamonte & D'Arcy, 2007), which has shown differences across dialects in terms of the development of these modals, thus making it necessary to limit this study to a specific region. Regrettably, the historic data available in Mexico are not confined to one region so the corpus consisting of literary works is representative of the country as a whole and not one particular dialect. This does not negate the number of different dialects that exist in Mexico, but simply exemplifies the dearth of regional documents available for a diachronic study (Aaron, 2007).

In order to capture both the development of these obligation markers through time and the reality of these constructions in modern Mexican Spanish three corpora were used: an oral corpus, a literary corpus and another small digitalized corpus comprised of written non-literary documents. Each corpus serves a unique purpose in this study, but as a group the three corpora will give a clear picture of the development of obligation in Mexican Spanish.

#### **2.1.1 Written Corpora**

In order to get a clear picture of grammaticalization, a diachronic perspective was necessary. The literary corpus includes Mexican literary prose from the 16th to the 20th

centuries.<sup>1</sup> The majority of these works were extracted from two digitalized corpora. The first, *Corpus Diacrónico del Español* (CORDE), has a total word count of more than 250 million words, but only works from Mexico were used. The Mexican section includes 7,822,486 words. This corpus is comprised of plays, prose, poetry, legal documents, historical accounts, religious works, and newspaper articles; however, only prose or historical works were used for this study.<sup>2</sup> The second corpus, *Corpus de Referencia del Español Actual* (CREA), includes the same range of written work as CORDE and

Table 2-1. Frequency count of each variant per century in the written corpora<sup>3</sup>

Century (word count)	TQ N	Normalized per 5,000	TD N	Norm	HD N	Norm	HQ N	Norm	Total
16 <sup>th</sup> (~2,339,244)	32	.11	89	.29	1091	3.58	21	.07	1233
17 <sup>th</sup> (~960,832)	40	.17	22	.10	852	3.70	19	.08	933
18 <sup>th</sup> (~403,340)	39	.50	12	.15	256	3.28	9	.12	316
19 <sup>th</sup> (~789,217)	253	1.00	2	.01	687	.67	79	.32	1021
20 <sup>th</sup> (~3,501,511)	865	2.41	3	.01	354	.99	232	.65	1454
Total N	1563		128		3248		381		4957

has more than 160 million words but; the total word count of the works from Mexico is 16,897,942. In addition, due to the lack of Mexican prose available in some of the earlier centuries, several works from the internet archive (<http://archive.org/details/texts>) were also included. Between the two corpora and the internet archive, and after eliminating the non-prose data, the total word count for the dataset in this study is 7,994,144. From

<sup>1</sup> The 16th and 17th centuries do contain some historical and legal documents owing to the general dearth of written documents available during this period. There was an effort, nevertheless, to use as few of these types of documents as possible.

<sup>2</sup> It is possible, although not expected, that these constructions are used differently depending on the genre. Consequently, the study was limited to prose so as to eliminate this possible limitation.

<sup>3</sup> The abbreviations found in the tables throughout this study represent each of the variants: TQ=*tener que*, HQ=*haber que*, TD=*tener de*, HD=*haber de*. Additionally, each variant was normalized per 5,000 words as a way to make the frequency counts across century unambiguous and are easier to compare.

these words, a total of 5,055 occurrences were extracted, as can be seen in Table 2-1.<sup>4</sup> For a complete list of texts used in this study, see Appendix A.

The other written corpus, *Documentos Linguisticos de la Nueva Espana: Altiplano Central* (Company Company, 1994; hereinafter DLNE), is relatively small and consists of 320 documents from the 15<sup>th</sup> through the 19<sup>th</sup> centuries and includes nearly 260,000 words (Aaron, 2004: 589). Most of these documents are letters or accounts of incidents recorded by a scribe. Schneider (2002: 76) states that documents of this type often ‘reflect many features of [...] speech fairly accurately’. The hope is that if there are differences among the use of obligation markers in formal (i.e. written) and vernacular speech, these two historical corpora (in addition to the synchronic analyses) will expose those differences. Every occurrence of the aforementioned obligation markers was extracted from this corpus and divided according to century.<sup>5</sup> Frequency counts of the examples found in the DLNE corpus can be seen in Table 2-2.

Table 2-2. Frequency count of each variant per century in the oral and DLNE corpora

Century (word count)	TQ N	Normalized per 5,000	TD N	Norm	HD N	Norm	HQ N	Norm	Total
16 <sup>th</sup> (~83,372)	8	.25	6	.19	108	3.42	6	.19	128
17 <sup>th</sup> (~85,321)	16	.47	5	.15	107	3.16	9	.27	137
18 <sup>th</sup> (~71,687)	27	1.26	6	.28	53	2.47	1	.05	87
19 <sup>th</sup> (~19,767)	4	1.01	--	--	11	2.79	1	.25	16
21 <sup>st</sup> spoken (189,840)	334	3.73	0	.00	8	.09	21	.23	363
Total N	389		17		287		38		734

### 2.1.2 Oral Corpus

The spoken corpus serves two purposes. First, as mentioned before, it can provide a clearer picture of the development of these obligation markers, particularly in

<sup>4</sup> The table only reflects a total of 4957 examples because 98 tokens were excluded from the written corpus, which is discussed in detail in section 2.2.

<sup>5</sup> Although, in general, data were divided up according to century, in some analyses the 17th and 18th centuries had to be combined given the lack of literature available during the 18th century.

vernacular speech. This is because, as Torres Cacoullos and Walker (2009) point out, the lexical history of a form has been empirically substantiated through patterns of use in the newer forms. Second, sociolinguistic interviews are used as a way to account for the possible sociolinguistic variables that may condition the use of each of the obligation markers. Including a corpus that allows for the analysis of sociolinguistic variables is based on the general Labovian idea that, in order

[t]o extract evidence for change in progress, we must separate the variation due to change from the variation due to social factors like sex, social class, social networks, and ethnicity, and from the variation due to internal factors like sentence stress, segmental environment, word order, and phrase structure (1994: 26).

Without sociolinguistic interviews, this separation would be impossible. The corpus is comprised of 26 sociolinguistic interviews with Mexican men and women from a small community in the state of Guanajuato, all of which were conducted in the summer of 2010. The total word count of this corpus is 189,840 and there are 333 occurrences of obligation found in this corpus (see Table 2-2). In order to obtain these data, however, two factors were taken into consideration: the speech community and participant selection.

### **2.1.2.1 The speech community**

The speech community used in this study is located in a small town in the state of Guanajuato, Mexico. This area was chosen (i) because of a longstanding friendship with families in the community and (ii) because this community could be considered a relic area. Poplack and Tagliamonte (2001) identify specific characteristics of a relic area: (i) lengthy settlement history<sup>6</sup> (ii) minimal immigration of non-members, (iii) historical

---

<sup>6</sup> Although Poplack and Tagliamonte indicate length being approximately two centuries, this community was not in existence two centuries ago. When the Agrarian Decree of January 1, 1915 passed, land was

continuity of current informants with the original settlers, (iv) geographic remoteness, (v) physical and psychological separation from adjacent populations and (vi) a strong sense of group identity (2001: 66). According to these criteria, this community qualifies as a relic area. This is particularly important given that relic areas commonly preserve older features of a language (Campbell, 1998; Poplack and Tagliamonte, 2001). This community is largely isolated located about two miles off a country road and more than ten miles away from the highway that leads to the two closest towns, Pénjamo and La Piedad, both of which are at least an hour's drive away. There are other communities nearby, although these too appear to be part of the larger speech community. However, no interviews were conducted outside this community.

#### **2.1.2.2 Participant selection**

Since all interviews were conducted in one summer, it was imperative that I enter the target community already having a connection i.e. friend of friend method (Milroy, 1987); the participants must feel comfortable during an interview in order to produce a natural variety of Spanish (Labov, 1994). The community has a population of about 300 people but my association with several members in the community made my acceptance unproblematic. As a result, it was relatively easy finding members of the community who wanted to participate. In addition to my connections, I also spent the first three weeks of my time there familiarizing myself with the community. Once I began to interview participants, an effort was made to consult an equal amount of men and women of a range of ages, socioeconomic backgrounds, and religions.<sup>7</sup>

---

distributed many Mexican families. It was during this time that the community was established. Prior to the establishment, most members of the community traveled around looking for work.

<sup>7</sup> It was thought that religion may be a possible conditioning factor given that Christian religions are often centered around what should and should not be done, according to the bible. Taking into consideration

The age of the participants range from 18 to 80 years and had all spent the majority of their lives in this particularly community. All were born in Mexico and Spanish is their first and only language. Most participants come from families who were part of the original settlement in the area, which happened shortly after the constitution of 1917.

The interviews were conducted with 26 residents, 14 women and 12 men using a digital recorder (see Appendix B for sample of interview questions). In an effort to minimize the effect of the OBSERVER'S PARADOX (Labov, 1972), the recorder was always placed in an inconspicuous place. Additionally, I did not ask participants to be 'interviewed'; rather I asked them if they would 'talk' with me about life in the community. The questions often invoked narratives about the participant's past, family, or general goings on in the community. Labov (1972) identifies these types of narratives as a way to take the participant's mind off the interview and elicit vernacular speech. Participants were also given the choice to be interviewed outside, in their homes or at the home of the family with whom I was staying. Crucially, these are all ways to minimize the effects of the observer's paradox (Labov, 1972).

Besides creating questions that invoked narratives, I was also careful to design my questions in a way that would educe expressions of obligation, like in 2.1.

- 2.1 *Tenía muchas responsabilidades en su niñez? Explica.*  
'Did you have a lot of responsibilities as a child? Explain.'

---

that obligation based on external forces (i.e. the word of God) is often deontic, it was thought that religious people may be use the obligation markers in different ways than their non-christian counterparts. This was unable to be measured, however, because the vast majority of the community was Christian.

This does not mean, however, that I was limited to these questions. Naturally, depending on the participant and the flow of the conversation, these questions varied greatly.

## 2.2 Data Collection

Once the interviews were conducted and transcribed, all occurrences of the aforementioned obligation markers were extracted from the three corpora. With the CORDE, CREA and DLNE corpora, this was done via computer programs like Microsoft Excel and Word that allow for a careful extraction of every token. With the literary works not from CORDE or CREA, however, an exhaustive search for every occurrence was done by hand. Once every token was accounted for, I then verified that each obligation marker was part of the variable context.

Owing to the tendency for these constructions to be polysemous, I excluded the following types of contexts: (i) cases where obligation/epistemicity were not expressed (2.2); (ii) false starts during the oral interview (2.3) or (iii) inability to decipher what the participant was saying (2.4). Table 2-3 shows the exclusions for *tener de* and *tener que* (T) and for *haber de* and *haber que* (Q) per century.

- 2.2 *Deseo **tengo de** verle (Calderón de la Barca, Pedro, 1634; CORDE)*  
'I have desire to see you/him/her.'
- 2.3 ***Teníamos que** pagar ci--, teníamos que pagar cinco a la semana o treinta al mes. (Roberto, 7/2010)*  
'We have to pay –we have to pay five per week or thirty per month.'
- 2.4 *Pues le gusta **haber que** xxx. (Reinaldo, 7/2010)*  
'Well, he likes to have to xxx.'

Once these tokens were excluded the total amount of occurrences was 5,691. This includes all written and oral corpora: 371 in the DLNE (Table 2-2), 363 in the oral corpus (Table 2-2) and 4957 in the written corpus (Table 2-1). These corpora will be

analyzed separately and the results of each analysis will be discussed and compared in Chapter 4.

Table 2-3. Exclusions of *tener que/de* (T) and *haber de/que* diachronically

	16 <sup>th</sup>		17 <sup>th</sup>		18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>		21 <sup>st</sup>	
	T	H	T	H	T	H	T	H	T	H	T	H
Expressed Possession	--27	4	42	3	6		6	3	4		--	
Lexicalized												
False starts	--	--	--	--	--	--	--	--	--	--	2	1
Difficult to decipher	--	--	--	--	--	--	--	--	--	--	2	2
Total	27	4	42	3	6		6	4	4		4	3

### 2.3 The Variable Context and Form-Function Asymmetry

A question one might ask is whether or not we can say that these forms are synonymous. That is, does each of these obligation markers essentially express the same thing? Linguists such as Lavandera (1978) seem to think they don't. However, from a variationist perspective it is recognized that different forms, like these obligation markers, can indeed be used for the same function (Tagliamonte, 2006). However, it is also important to acknowledge that each form likely has a range of functions, especially if they are undergoing change. There may only be overlap with a limited number of the functions and it is therefore important to define the context(s) where these functions overlap, since it is in these contexts that the variants co-occur and the patterns of use of the forms can be discovered.

Torres Cacoullos (2001) demonstrates this in her study on the grammaticalization of Spanish *estar* + gerund and *andar* + gerund. She illustrates how variation studies can in fact explain the patterns of use of these two forms, despite the asymmetry of form and function. She goes on to explain that, based on the different histories of each of these constructions, there will always be contexts in which there is no overlap but, as they grammaticalize, they move towards becoming variants of the same variable. When

they do start competing for the same territory, the space is not necessarily homogenous, given the different histories of the forms (Aaron, 2010; Torres Cacoullos, 2001). However, according to the NEUTRALIZATION-IN-DISOURSE HYPOTHESIS, the meaning nuances that distinguish each form may not always be considered by the speaker when making a choice, particularly in this shared space (Sankoff, 1988). We can assume then that obligation is no different and that, even though there are overlapping functions, each form likely has a bigger inventory of distinct functions.

In the case of Spanish obligation, however, there have been limited studies that have focused on the semantic differences between these forms (c.f. Rabadan, 2006) Of the little research that has discussed these differences, very few have empirically backed assumptions, and instead rely on intuition as a way to describe these differences (e.g. De Maeseneer, 1998). Using a variationist approach minimizes dependence on intuitions and, instead, operationalizes the possible differences in meaning among the forms.

#### **2.4 Circumscribing the Variable Context**

When trying to circumscribe the variable context, it is necessary to include 'the largest environment in which this variation occurs, in order to apply the PRINCIPLE OF ACCOUNTABILITY,' which states that every occurrence needs to be accounted for, including both the environments where the form occurs and where it could occur (Labov, 2008: 2) The variable context should additionally exclude instances where one of the linguistic variables cannot be substituted with another (Poplack and Tagliamonte, 2001). This allows for a clear analysis of the dependent variable and not the various other functions that these forms may have (Labov, 1972: 72). Similar to defining the variable context, not accounting for every possible occurrence can yield undependable

results. Applying the principle of accountability, however, is especially difficult in morphosyntactic studies, as it is nearly impossible to identify and record all of the contexts where the dependent variable occurs and could occur. With the application of quantitative sociolinguistic analyses to morphosyntactic studies many linguists have argued for the validity of this type of analysis with non-phonological variables (e.g. Labov, 1982; Winford 1993). Winford, for example, points out the need to use the principle of accountability in a way that still adheres to finding variables in the same context but not requiring every occurrence or non-occurrence to be considered, given its near impossibility (Winford, 1993).

For the purposes of this study, the variable context is obligation, as expressed by the aforementioned constructions. The variable context includes all the occurrences of these obligation constructions in the overlapping functional space. Furthermore, it will necessarily exclude instances of the same constructions in contexts where one of the other constructions cannot be substituted or when obligation was not expressed. Thus, a construction like *tener que* would only comprise contexts where this form had an obligation meaning and not meanings such as possession. (see section 2.2 for additional exclusions). Once the envelope of variation has been established, a diachronic perspective of the development of these forms will be examined. This will be done in part using a variable rule analysis, which will be discussed in the next section.

## **2.5 Variable Rule Analysis**

Taking into consideration the three corpora used in this study there are 5691 occurrences of obligation markers which is an adequate amount of tokens for a multivariate analysis (Guy, 1988; Bayley, 2002). Provided that the variable context has been clearly defined (see sections 2.3 and 2.4), factors thought to condition the use of

each of these markers were chosen. Each of these factors was based on previous research and will be discussed in detail in Sections 2.6.1-2.6.10.

Once the factors were chosen each of these tokens was then coded in regards to its occurrence with the aforementioned markers in an excel spreadsheet (see Appendix B for list of factor groups and factors). Once all the data were coded they were analyzed quantitatively. For this process I used GoldVarb (Cedergren and Sankoff, 1974, Sankoff, 1988), a statistical multivariate analysis program that takes the idea of structured heterogeneity (Weinreich et al., 1968) and quantifies it (Tagliamonte, 2006). This allows for probabilistic predictions of the conditioning effects of each of these obligation markers. That is, we can examine from a quantitative perspective the surrounding linguistic context to identify which variables (linguistic or extralinguistic) condition each obligation marker (Bayley, 2002).

The program also provides frequency distributions of each variable relative to the various factor groups as well as the individual factors thus providing a well-defined representation of the semantic environment of each of the obligation constructions. Importantly, these analyses do not rely solely on surface similarities. Rather this approach takes into account that sometimes forms that appear to have similar functions on the surface do not necessarily signify that they are similar. Similarities are determined through a variable rule analysis, where the comparison of conditioning effects across time (real or apparent) is considered. This is done through the examinations of the HIERARCHY OF CONSTRAINTS, or the order of effect that the independent variables have on the dependent variable. This type of analysis allows for the measurement of various factors at one time and orders them in terms of the effect

they have on the dependent variable thereby giving a fine-grained representation of the system under examination and a basis with which to compare. This is in contrast to surface comparisons whereby assumptions are made that likeness in form denotes similarity in function.

Multivariate analysis allows me to test for each of the hypotheses (discussed in detail in sections 2.6.1-2.6.10), as well as see the general direction of change of each of these obligation constructions by comparing the results (hierarchy of effect) of these analyses across centuries or age groups (see section 2.6 for a discussion on how this is done). For example, with the factor group grammatical person and animacy, the results will first show whether the factor group is significant. If it is significant, we can then look at the probability values of each factor. These values will tell us if this significant factor favors (greater than .50) the obligation marker, or disfavors it (lower than .50) essentially showing the strength of association of each factor with the form (Bayley, 2002).

If a factor group is not significant there are still conclusions that can be obtained. For example, if age is not a significant factor group, then we can postulate a variety of reasons why this is the case: (i) that these constructions undergo such gradual change that differences are too slow to be apparent between generations; (ii) that perhaps the age division was incorrectly defined and the insignificance is due to this; or (iii) we could also look in the data for trends to see if there may be tendencies among age groups and draw a tentative conclusion. Crucially, significant results are important, but of equal importance are the conclusions we extract from the results.

Results should corroborate or refute prior conclusions drawn in the literature and essentially position the findings in terms of theory. As Bayley (2002: 130) states ‘explanations must be sought in linguistic theory and in our understanding of the history and social structure of the communities we study.’ In cases of grammaticalization, the comparative sociolinguistic method is a technique that situates the results from a variable rule analysis in a way that clear conclusions can be drawn about the development of each form and the changes the forms incurred throughout history.

## **2.6 Comparative Sociolinguistic Method**

The comparative sociolinguistic method is a technique that has been adopted from diachronic historical linguistics as a way to examine similarities between dialects or varieties synchronically. In contrast to historical linguistics, these analyses do not rely solely on surface similarities. Rather this approach takes into account that sometimes forms that appear to have similar functions on the surface do not necessarily signify that they are similar. Similarities are determined through a variable rule analysis where the comparison of conditioning effects (factors) across time (real or apparent) or across dialects is considered (e.g. Poplack & Tagliamonte, 2001; Tagliamonte, 2002; Tagliamonte, Smith & Lawrence, 2005; Tagliamonte & Smith, 2006; Tagliamonte, 2006a, 2006b; Schwenter & Torres Cacoullos, 2008).

One of the core principles of the comparative method is the UNIFORMITARIAN PRINCIPLE, which states that the factors that effected language change in past centuries are likely to be similar to the factors that prompt language change in modern languages (Poplack & Tagliamonte, 2001 and references therein). Additionally, Labov (1972) indicates that synchronic variation is often a reflection of diachronic change so detecting changes in a modern system is essentially a window into the past. It has been pointed

out in the literature that forms often retain connotations that are related to their original meaning (e.g. Bybee et al. 1994; Poplack and Tagliamonte, 2001). Thus a synchronic study of a linguistic variable is likely to reveal certain patterns of use and semantic properties that can be connected to the past. Fundamentally, reconstructing the language with synchronic principles complements the data already available by offering a new way to construct the history and development of a language.

An important component of the comparative sociolinguistic method is looking for correspondences in the variable context (Poplack and Tagliamonte 2001: 96). These correspondences are in reference to similarities found in the order of effect of the conditioning factors in two (or more) sources. That is, once the variable rule analysis has provided significant results the order of conditioning effects between the sources – in this case the source is represented by the centuries-- is compared (Tagliamonte, 2001: 732).

When similarities are found in the conditioning effects of the forms we can deduce that these two sources are related, whereas in contrast, if no similarities are found, it is possible that the two sources are not related (Tagliamonte, 2002 and references therein). An example of this is in a study conducted by Poplack and Tagliamonte (2001) on five dialects of English, three of which are posited to be early varieties of African American English. In these three varieties they found identical conditioning effects thereby confirming their belief that these dialects derive from a similar source (2001: 232). In the present study, we know that the source language is the same, so this step is less informative in terms of results.

Another important element of the comparative sociolinguistic method is the idea of a CONFLICT SITE (Poplack and Tagliamonte, 2001). This is defined as 'a form or class of forms which differs functionally and/or structurally and/or quantitatively across the varieties in question' (Poplack and Meechan, 1998: 132). In other words, once the variable rule analysis has been performed and correspondences have been verified, the conditioning effects of each source are compared. If there are differences in the order of the factors within each significant factor group then this is considered a conflict site. This is particularly important for the present study since these conflict sites are indications of changes occurring in the system. These conflict sites will help identify where on the cline of grammaticalization these constructions may be in each century (Poplack and Tagliamonte, 2001; Tagliamonte, 2002). Additionally, these conflict sites may pinpoint where in the grammar these changes are realized.

There have been several variationist studies on modality that have employed the comparative sociolinguistic method. Tagliamonte (2004) uses this method in an apparent time study she conducts on the modal system of Northern British English. Her results reveal several patterns, such as the diminishing use of *must* across age groups (p. 51). She also concludes that, in contrast to other dialects spoken in the south, *have to* and *have got to* are still relatively stable in this variety, citing specialized functions as the primary reason for this distinction (2004: 51). These results are compared to other studies conducted on modality in other dialects such as Tagliamonte and Smith (2006) and Tagliamonte and D'Arcy (2007).

When comparing Tagliamonte and Smith's (2006) study on English in Northern England, similar results are found with the modal *must*, which indicates a general move

of the deontic modal towards obsolescence in the British dialects. They also find that *have to* appears in contexts that were historically reserved for *must* while *got to* is associated with contexts that have indefinite reference (Tagliamonte & Smith, 2006:342). Compare these results to the modal system in Toronto English where Tagliamonte and D'Arcy (2007) conclude that *have to* has essentially taken over the entire modal system. They claim that this is a strong indication that English modality is in flux and it appears as though the Canadian English system has undergone the most change of the three varieties (Tagliamonte & D'Arcy, 2007: 82).

The general lack of research on modality from a variationist perspective clearly illustrates the need for more studies like this. Applying the comparative method will provide a baseline for research on modality in other languages by showing us the path of change (e.g. Tagliamonte & Smith, 2006). If other studies mirror the paths found in this study, we can describe this path in more detail, perhaps including where in the grammar these changes begin to take place.

## **2.7 Variables and Hypotheses**

In a usage-based framework, variation in language is viewed as structured and ordered, yet heterogeneous as put forth in the principle of structured heterogeneity (Weinreich, Labov & Herzog, 1968). In order to study and analyze this variation in a reliable way, Labov (1972, 1994, 2004) has outlined certain fundamentals. One of these is the identification of the independent and dependent variables. The dependent variable is the form or forms that are the focus of the study whereas the independent variables are the factors that may or may not have a conditioning effect on the use of the dependent variable (Labov 1972, 2004).

The independent variables chosen for this study as possible conditioning factors were selected either because i) previous research on modality has identified them as pertinent or ii) research on universal paths of grammaticalization has categorized them as possible indicators of a grammaticalized form. The independent variables that will be considered in this study are: type of verb, animacy, type of sentence, polarity, objects (direct, indirect), preceding clitic, tense-aspect-mood, temporal adverbs, lexical verb type, sex and age (see Appendix C for a complete list). In sections 2.6.1 to 2.6.10, I will discuss the independent variables chosen for this study, as well as the hypotheses supporting each of these.

### **2.7.1 Verb Type**

Verb type is a common conditioning factor in morphosyntactic variation (e.g. Tagliamonte & Poplack, 1993; Silva-Corvalán, 1994; Casanova, 1999; Mayoral Hernández, 2007; Cornillie, 2007) and therefore often considered in these types of studies. For the present study, verb type refers to the classification of the infinitive verb that follows that follows the obligation marker. Many of these classifications are based on a study conducted by Cornillie (2007).

Cornillie (2007: 219), who considers verb type in his study on evidential and epistemic modality in Spanish, looks at the frequency distribution of the deontic and epistemic uses of *tener que* and finds that it is overwhelmingly used as a deontic modal with action verbs (99.5%). He finds similar results with verbs of perception (100%) and mental activity verbs (96.1%). In fact, the only verb type that falls below 90% range for deontic modals is the type Cornillie defines as copular verbs and other verbs referring to attribution (85.12%). This result implies that these copular verbs and verbs of attribution may be the access point for expressing epistemicity for *tener que*. A tendency for

copular verbs to appear with epistemic modals will further support studies such as Pietrandrea (2005) and Coates (1983) who find that stative (e.g. copular) verbs are often found in epistemic contexts.

It is important to recognize that Cornillie also had a verb type for mental activity and perception, which can easily overlap with verbs of attribution. Thus for the present study, verbs of attribution will be divided into separate clearly defined groups: verbs of perception, psychological/emotion verbs, copular verbs and communication verbs. If the statistics show similarities among these three factors, they will be collapsed into one group.

Besides Cornillie (2007), verb type has not been considered as a possible factor in determining the use of one variant over another in other studies on obligation markers. I posit that verb type is crucial to this process, given that these collocations are dependent on the presence of a nonfinite verb. Therefore, it is suspected that these non-finite verbs are partially responsible in determining which variant will be employed. Owing to the frequency distribution of *tener que* in Cornillie, it is postulated that this construction is more likely to occur with actions verbs early on, but as it moves towards epistemicity, it should begin to appear more often with copular verbs and other stative verbs. Taking into consideration that *haber de* is so frequent in the earlier centuries, it may be that it is favored by the verb types that encompass attribution early on but, as *tener que* increases in frequency, this may diminish.

### **2.7.2 Subject and Animacy**

Another common conditioning factor found in many variationist studies (e.g. Tagliamonte & Poplack, 1993; Tagliamonte & D'Arcy, 2007; Cornillie, 2007; Szmrecsanyi & Hinrichs, 2008; Meyerhoff, 2009) is grammatical person and animacy.

This is particularly important when the variable under examination is a verbal construction, given that the subject is intricately bound to the verb.

In relation to obligation markers, Cornillie (2007: 219) measured the frequency of the different grammatical subjects and animacy with *tener que* comparing its use in epistemic and deontic contexts with each type of grammatical person (including animate and inanimate). In his data he finds that deontic *tener que* occurs with animate subjects 97.75% of the time. This is not surprising given that the subject of obligation markers is frequently animate (Tagliamonte, 2002). In addition, Coates (1983: 36) states that 'it is generally true that examples with second person subjects are stronger (in terms of obligation) than those with first person subjects, while examples with first-person subjects are usually stronger than those with third person subjects.

Building on this, I postulate that as *tener que* advances along the grammaticalization path towards epistemicity it will be less and less favored by animate subjects, particularly in first and second persons since these are strongly associated with strong obligation. I expect that animacy, from a diachronic perspective, is an area where change occurs. This is based on patterns found in studies by linguists such as Coates (1983), Bybee and Fleischman (1995) and Pietrandrea (2005), who all suggest that contexts expressing deonticity occur more often in first and second person subjects, whereas epistemic contexts are more likely to occur with third person subjects. Research also correlates animacy with high focus clauses, which tend to be less grammaticalized than their inanimate, low-focus counterparts (Klein Andreu, 1991; Bybee et al., 1994; Diaz-Campos, 2011).<sup>8</sup>

---

<sup>8</sup> Klein-Andreu (1991) relates low-focus contexts to relative and subordinate clauses whereas the high-focus clauses are more likely to be main clauses. Other literature (e.g. Bybee et al. 1994) attributes low

Taking this into consideration I postulate that, during the 15th and 16th centuries, it is likely that animate subjects will favor *tener que*, predominantly first and second person animate subjects. This would mean that, as these constructions develop and move into epistemic contexts, they should shift from contexts with first and second person subjects into contexts with third person subjects.

However, given that frequency plays an important part in this change, and *tener que* suffers the most dramatic shift in frequency across centuries, this construction will most likely be the one to undergo this shift in subject preference. That is, in the later centuries I believe that *tener que* will begin to also be favored by third person as it moves along the grammatical cline while its occurrence with animate subjects becomes less marked.

Given the apparent advanced stage of the collocation *haber de*, it is likely that this construction is already favored by third person in the early centuries. In the later centuries, as *tener que* increases in frequency and starts to take over the territory of the other forms, we may see *haber de* lose ground in this area.

### **2.7.3 Sentence Type**

Declarative or Interrogative sentences are the two sentence types that are included in the present study. Like the previous factors, sentence type is a factor that has been considered in a large number of variationist studies (e.g. Smith, Durham and Fortune, 2007, Cornillie, 2007; Tagliamonte and D'Arcy, 2007; Alamillo, 2009; Torres

---

focus to clauses that provide background information and high focus to clauses that provide new information. For the purpose of this study, these clauses will be identified primarily by their tendency to occur with the aforementioned factors (e.g. negation, stative verbs, (in)animate subjects)

Cacoullos, 2011). Although this factor group is often a conditioning factor in these studies, it has only been considered in a limited number of studies on modality.

In a study conducted on Toronto English obligation markers Tagliamonte and D'Arcy (2007:61) found that *have to* occurred almost categorically in interrogative sentences. As a result they were unable to include it in the multivariate analysis. Nevertheless, they indicate interrogative contexts as a place of interest although they do not expound on why this is. Furthermore, Coates (1983: 242) also observes that epistemic modals generally do not occur with interrogative constructions. Taking into consideration that these epistemic contexts suggest a high degree of certainty in relation to the proposition, and an interrogative sentence would likely cast a degree of doubt on the proposition, this could be an explanation as to why these modals occur less often in interrogative sentences when expressing epistemicity.

Whether this is the case or not, we can postulate that, based on the frequency of *haber de* in the earlier centuries and the frequency of *tener que* in the later centuries, these constructions are less likely to occur in interrogative sentences at different points in history. In the earlier centuries, *haber de* will likely have a disfavoring effect with interrogative sentences, but as *tener que* increases in frequency and likely continues to grammaticalize this disfavoring effect may shift towards *tener que*.

In addition, the movement of *haber de* into epistemic contexts could be an indication that the newer form, in this case *tener que*, is grammaticalizing. Aaron (2010: 30) indicates that when an older construction generalizes semantically, this can facilitate the movement of newer forms into the former territory of the older form. In this case, if *haber de* has moved into epistemic contexts, it will not be used with interrogatives in

those contexts, thus giving *tener que* access to these types of sentences.

Consequently, in the earlier centuries, I posit that interrogative sentences will disfavor *haber de* while at the same time favoring *tener que* and the other variants. But as *tener que* expands in the later centuries the other variant (*haber que*) will likely be used more often in interrogative sentences.

#### **2.7.4 Polarity**

Polarity, which determines whether an item in an utterance is positive or negative, has long been considered to be a factor in linguistic change (e.g. Bolinger, 1977; Coates, 1983; Aaron & Torres Cacoullous, 2005; Tagliamonte & D’Arcy, 2007; Torres Cacoullous, 2011). In terms of modality, negation also seems to play an important role. Many of the studies that examine polarity in changing modal systems look at the placement of the negation in relation to the modal marker (e.g. Plank, 1984; Fernández de Castro, 1999). This, however, is not relevant to the present study, as the placement of the negation marker is fixed after the 16<sup>th</sup> century, occurring directly before the construction (or the clitics).

Negation is also mentioned when examining epistemicity. Coates (1983), for example, finds that, like with interrogative sentences, negation in English is limited in terms of its occurrence with epistemic contexts. This limitation on negation in epistemic contexts is echoed in research conducted on Spanish modals by Solano-Araya (1982: 17), although he is unclear about which epistemic modals can or cannot be negated. Coates (1983) and Plank (1984: 331) expand on this claim by illustrating that although English *must* is limited in negative epistemic contexts, it can be replaced with the modal *can’t*. This suppletive relationship between two epistemic modals has not mentioned in

the literature on Spanish modals despite claims of the limitations on negation and epistemic expressions.

Given these results, I also consider polarity as a factor group, postulating that contexts where negation occurs are likely to favor *tener que* early. As this construction increases in frequency and likely advances into more epistemic contexts in the later centuries, there should be a shift towards non-negated contexts. For the purpose of this study, an example was considered to be negated if the obligation marker or the infinitive verb that followed was negated, as in (2.5). Otherwise the example was coded as affirmative as in (2.6)

2.5 *yo no **tuve que** gritarles a ustedes* (Fuentes, 1962; CORDE)  
'I didn't have to yell at you guys.'

2.6 *¿Qué les **hemos de** hacer?*(Mendieta, 1604; CORDE)  
'What do we have to do to them?'

### 2.7.5 Objects

Although none of the previous research on modality has included objects in their analyses, it has been examined in other areas. Much of this research has focused on phenomena such as *leísmo* (e.g. Choi, 1998; Tippits, 2011) or clitic climbing (e.g. Myhill, 1988). These studies tend to examine the role of objects in terms of the placement of the object syntactically (which will be discussed in section 2.3.6) or whether the object is animate, inanimate, oblique etc., and the role these objects have in the grammaticalization process.

Research such as Klein Andreu (1991), Bybee et al. (1994), and Diaz Campos (2011) relates objects to the grammaticalization process by suggesting that, as certain constructions grammaticalize, they move from high focus clauses to low focus clauses. They correlate low focus clauses with more grammaticalized forms and identify these

clauses through their association with negation, inanimate objects or subjects and stative verbs. Although the focus of these aforementioned studies was not modality, it is still possible that a similar tendency is found in these constructions as they grammaticalize.

In addition to previous research, it is also important to take the uniformitarian principle into account (see section 2.6), which states that the factors that effected language change in past centuries are likely to be similar to the factors that prompt language change in modern languages (Poplack & Tagliamonte, 2001). Consequently, a study conducted on the Medieval Spanish possession markers *tener* and *haber* may provide an answer to the development of these obligation markers.

Urrutia and Alvarez (1995:212) examine the variation of these two markers and find that *haber* is more often utilized with abstract objects, whereas *tener* appears more with concrete objects. As previously mentioned, these tendencies may have continued as these possessive verbs developed into obligation markers, so this is a pertinent factor group for consideration. If this is the case a similar trend should be found in my data, whereby abstract objects are found with *haber de* and *haber que*, while *tener que* and *tener de* will be the default forms occurring in all other contexts.

As a result of the aforementioned studies, the animacy of these objects is considered as well as whether or not these objects are abstract or concrete. It is postulated that as each construction grammaticalizes and moves towards epistemicity, it will start to appear more often with inanimate objects essentially becoming more and more associated with low focus clauses. Given the high frequency of the construction *haber de* in the earlier centuries, I posit that contexts with animate objects are more

likely to be areas in the grammar for the newer constructions, like *tener que*, *haber que* and *tener de*, since these type of contexts are associated with high focus clauses. Conversely, I postulate that as these constructions grammaticalize, particularly *tener que*, they will more likely occur in contexts with inanimate objects.

For this study, I considered whether animate (2.7), inanimate objects (2.8), obliques (2.9) or the middle voice (2.10) play a role in the choice of these constructions. In addition, objects were divided into two groups: direct and indirect objects.

- 2.7    ¿no te **he de** amar? (Fernández de Lizardi, 1818; CORDE)  
         'I don't have to love you?'
- 2.8    **tuvo que** atravesar el río (Heriberto, 1893; Internet Archive)  
         'he had to cross the river'
- 2.9    *aquello* **tenía que** acabar á las doce (Riva Palacio, 1896; CORDE)  
         'That one had to end at twelve o'clock.'
- 2.10   **tenían que** retirarse a sus casas. (Fernández de Lizardi, 1818; CORDE)  
         'They had to withdraw to their houses'

### 2.7.6 Preceding Clitic

The next linguistic factor to be considered in the study is preceding clitic. Similar to the previous factor group, preceding clitic has not been considered in previous research on modality. Myhill (1998), however, looks at clitics and the semantic differences between preceding and attached clitics and concludes that the position of a clitic largely depends on the status of the construction as a lexical or auxiliary verb. Given that these obligation markers are grammaticalizing and auxiliarihood is in fact a part of this process, it is possible that the form that has moved furthest along the grammaticalization path i.e. *haber de* is likely to be favored by preceding clitics. This is assuming that the construction is being processed as a unit, so the clitic is attaching to the infinitive based on the status of the obligation marker (as a semi-auxiliary) plus

infinitive.<sup>9</sup> Any clitic that is not attached to the infinitive will be considered a preceding clitic. This includes clitics that come directly before the infinitive or attached to the obligation marker (2.11) or before the obligation marker (2.12).

2.11 *¿hanse de perder todo esto?* (Anónimo, 1541; CORDE)  
'Do you (pl) have to lose all of this?'

2.12 *se le había de morir algún hijo o hija* (Toribio de Benavente, 1536 – 1541; CORDE)  
'one of his children had to have died.'

### 2.7.7 Temporal Adverbs

Considering that the most thorough research conducted on modality has been limited to the English modal system and that the approach taken by Bybee et al.(1994) is more focused on identifying cross-linguistic tendencies, it is possible that there are developments within agent oriented modality that have not yet been substantiated. One possible inclination that has not been included in the research is the habitual reference of *haber que* and *tener que* in such examples as (2.13-2.14). Coates (1983: 54) mentions it in her study of English modal auxiliaries pointing out that *have to* can be habitual in meaning while modals like *have got to* and *must* cannot.

2.13 *porque todos los días habíamos de confesarnos* (González, 1999; CREA)  
'because every day we had to confess.'

---

<sup>9</sup> Many studies have emphasized the difference between modal auxiliaries and periphrastics, highlighting that modal auxiliaries are more susceptible to change (Coates, 1983; Westney, 1995). For example, Westney (1995:31) claims that modal auxiliaries are more likely to grammaticalize stating that the 'periphrastic items tend to manifest epistemic senses only to a relatively restricted degree'. In fact, there has been a tendency to separate true modal auxiliaries like 'must' and 'can' from quasi-modals like 'have to' and 'be going to' in order to compare the respective progression of these constructions. Consequently, criteria for determining the status of a modal verb as auxiliary or quasi-modal i.e. periphrastic have been established, most notably Palmer's 'modal criteria' (Westney, 1995, and references therein). If a verb does not meet these criteria then it is not a modal auxiliary but rather a periphrastic or lexical verb (Westney, 1995;15). Considering that these criteria are language specific to English it is questionable whether the claim Westney makes is valid or not in Spanish. Criteria for determining modal auxiliaries in English have been dubbed NICE properties which define the modal auxiliaries: negation, inversion, code, and emphatic affirmation. For a more in depth discussion on English modal auxiliaries see Coates (1983), Palmer (1987), Westney (1995) or Krug (2000).

- 2.14 *Todos los días tengo que meterme por ahí, a bailar.* (Fuentes, 1958; CORDE)  
'Every day I have to go there to dance.'

Westney (1995) also finds examples of habitual reference in his study of English modals. He correlates habitual action with future claiming that *have to* can be used habitually [...] to the extent that the repeated actions [...] extend before and after the present while with *have got to* and *must*, the repeated actions are restricted to the future (1995; 134). The only mention of habitual action in Bybee et al. (1994) is in reference to the past habitual use of *would*. They claim that its use as a habitual past is a separate development from the modal *would*. Crucially, they allude to the fact that this may be a possible lexical split for the construction *would* (1994; 238). Taking this into consideration, and the fact that we see examples of habitual reference with these obligation constructions, it is possible that this is a common development of certain agent-oriented constructions. Examples such as these should be acknowledged and examined in more detail.

In addition to habitual action, temporal adverbs may also be able to account for movement of these obligation markers into epistemic contexts. For example, Blas Arroyo (2008) found that the synthetic future, which is derived from an obligation marker (verb + *habere*), is favored by non-specific temporal adverbs. Although his study focuses on the future, because the synthetic future comes from an obligation marker and obligation, like future, grammaticalizes into epistemic contexts it is possible that temporal adverbs can elucidate this movement into epistemicity.

For the purpose of this study a temporal adverb will be any adverb that makes reference to time (2.15) but each temporal adverb will then be classified as specific or

non-specific (for studies that examine the effect of specific and non-specific adverbial markers, see Urrea & Gradoville, 2006; Aaron, 2006; Howe & Schwenter, 2008; Blas-Arroyo, 2008) so as to elucidate possible tendencies of habitual action, as in 2.14 above. Since there hasn't been any previous research on modality that has included temporal adverbs as a possible conditioning factor, my hypothesis is centered on Blas Arroyo's finding (2008).

2.15 *para el viaje que **tienes que** hacer mañana* (Ibargüengoitia, 1979; CORDE)  
'for the trip you have to take tomorrow'

Aaron (2006) and Torres and Walker (2008) find that as a form moves into more epistemic contexts it is more likely to occur with temporal adverbs in future contexts, which they suggest helps to clearly distinguish epistemic meanings from the future meaning. I would postulate then, that a similar process is occurring with obligation. That is, as these constructions move away from obligation and intention and into purely epistemic contexts they will less likely occur with temporal adverbs thus creating a clear distinction between the two functional spaces. Due to the limited number of adverbs present in the data these will not be included in the multivariate analysis but they will be discussed in detail in Chapter 4.

### **2.7.8 Tense, Aspect, Mood**

Another way to determine whether habitual action is part of the development of obligation is to look at aspect. However, in order to accurately code for aspect and avoid interaction with temporal adverbs, it was necessary to include mood and tense in this factor group.

Aside from habitual actions, tense may also play an important role in determining where these forms are in their development. Tagliamonte and D'Arcy (2007), for

example, look at obligation in Toronto English and find that all non-present tenses were categorically used with *have to*. In their study, it was determined that *have to* had, in essence, taken over the function of obligation. If this is the case with one obligation system, is it possible that tense has an effect on other systems? If this is the case and *haber de* is the more grammaticalized form and has moved towards more epistemic contexts, which of these forms has taken over? In her comparison of obligation marker use in several different canonical literary works De Maeseneer (1998) finds that *tener que* is the most frequent form used in Mexican Spanish.<sup>10</sup> Thus, it is likely that all non-present forms will favor the use of *tener que*.

### 2.7.9 Lexical Verb Type

The lexical verb that follows the obligation construction is also considered as a factor group. Similar to some of the other factor groups discussed, lexical verb type has not been considered in any studies that have focused on modality. It has been, however, the focus of many variationist studies. Torres Cacoullos and Walker (2008: 5) point out that lexical type is of particular importance when looking at the grammaticalization of a collocation since these are often the source of language change in diachrony. They conclude that lexical type plays an important role in determining the distribution patterns of the English future. In addition, they illustrate that these collocations occur with certain lexical verbs and eventually become fixed expressions, ultimately contributing to the variation of these future forms.

---

<sup>10</sup> She looked at two Latin American works, *El general en su laberinto* by Gabriel García Márquez and *Como agua para chocolate* by Laura Esquivel. She broke down the frequency counts of the various obligation markers in each of the works and *tener que* (n=124) occurred more seven times more often than the next most frequent marker, *haber que* (n=16)

Poplack (2001: 411) looks at lexical type in her study on the French subjunctive and finds that it is constrained when co-occurring with the subjunctive. Smith (2000) also finds that lexical verb type can have an effect on the variant in her study of negative *do* in Buckie Scots. Taking these previous studies into consideration, we can assume that lexical type may have an effect on one or more of the variants, whether it is through lexicalized expressions or highly constrained contexts. I would posit, like in verb type, that stative verbs, such as *ser* and *estar* 'to be' and *tener* 'to have', are more likely to be favored by *haber de* in the later centuries, assuming that this form has moved into primarily epistemic contexts. Although I expect to find a similar tendency among all stative verbs, there were enough examples of obligation markers occurring with *ser*, *estar* and *tener* to warrant separating these verbs from the rest of the stative verbs. Additionally, Pietrandrea (2005) found that stative verbs often occur with epistemic future in Italian, particularly the verbs *avere* 'to have' and *essere* 'to be'.

It is also postulated that in the earlier centuries, when *tener que* first comes into use as an obligation marker, it is not likely to be found with stative verbs (or copular verbs). This is because the form has just entered the functional space of obligation and may still have retention of the older form, possession. Therefore it may be limited to contexts in which both an obligation and a possession meaning can be gleaned thus limiting the lexical verb type with which it can appear. This is based on research conducted on the English obligation system and postulations that the construction *have to* as an obligation marker is derived from contexts such as *I have a secret to tell you*>*I have to tell you a secret* (Brinton, 1991; van der Gaaf, 1931).

I also postulate that in the later centuries as *tener que* continues to grammaticalize, it is found with a greater variety of lexical types than *hay que* or *haber de*. We see evidence of the expansion of *tener que* in Cornillie (2007: 219), who found that all verb types except copular verbs and verbs of attrition were used with *tener que* more than 90% of the time. *Hay que*, although also a quite frequent variant according to De Maeseneer (1998), lacks the liberty to appear in all grammatical persons, thus possibly limiting the number of verbs it can appear with.

The results and tendencies found in this factor group will be discussed in Chapter 4 but this factor group will not be included in the variable analysis due to interaction with verb type.

#### **2.7.10 Extralinguistic Variables**

As mentioned before data were collected on the age, sex, social class, and religious affiliation of each participant. The intention was to include each of these factors in the multivariate analysis. However, this is not possible with all of these categories. With religion, for example, due to the high number of Catholics in the community and the contrastive low number of other religions (n=2) represented in the speech sample, there were not enough participants to consider religion as a factor in the multivariate analysis.

A similar situation occurred with social class. Upon entering the community, I assumed that the majority of the members seemed to belong to a lower social class. However, after a couple of weeks, it was evident that there was stratification in the community via land ownership. According to Ash (2002:403) ‘two central components of social class are (i) the objective, economic measures of property ownership and the power and control it confers on its possessor, and (ii) the subjective measures of

prestige, reputation, and status'. In this community property ownership played a central role in status in the community. The property owners are in a power position, given that the rest of the community depends on them for work during harvest months. This consequently creates a natural class boundary.<sup>11</sup> In order to consider this factor in the analyses, however, there would have been several participants from each class. This was not possible due to the limited number of land owners who agreed to participate in the interviews. As a result, only two social variables were considered: age and sex.

#### **2.7.10.1 Age.**

Age is an important variable in order to capture any change in progress in apparent time. Indeed, apparent time is dependent on this dimension, where the element of time is created through the formation of age groups (Labov, 1972: 163). Although it is possible that distinctions among the different age groups could be due to age grading, i.e. linguistic changes that regularly occur during a certain period in life, these claims have rarely been substantiated (Bailey, 2002 and references therein). An emic approach was used in order to divide the age groups (Eckert, 1997). That is, I subjectively devised groups in terms of life stages as a means to group speakers according to age. This resulted in two different age groups; young adulthood, and adulthood. This division seemed to be a natural division since most of the adults older than 45 years seemed to congregate together whereas the adults younger than 45 years congregated in separate areas of the community.<sup>12</sup> Although it would have been

---

<sup>11</sup> There is one exception to this group. One female, who does not own land, is still considered to be in the middle lower class because she receives money every month from the U.S. government for the death of her husband. This makes her less dependent on the land owners, and neutralizes the power relationship.

<sup>12</sup> Of course there was also a group of people under the age of 18 that also congregated but I did not interview anyone younger than 18 so this group was not included in the analysis.

advantageous to have another group, possibly separating middle-aged participants from older participants due to life experience, there were not enough female and male participants to do this.<sup>13</sup>

In regards to the use of the obligation markers, it is posited that the younger generation will use *tener que* more given its rise in frequency across centuries. Additionally, I expect them to use *haber de* less often since Fernández de Castro (1999) suggests that this construction has moved into formal and written contexts. I postulate that the younger generation will also use one (or more) of the obligation constructions in more contexts than the older generation. However, Tagliamonte (2002) showed no significant effect from age and sex in her study of obligation markers in northern British English so it is also possible that age has no effect on variant choice.

#### **2.7.10.2 Sex**

The other extralinguistic factor examined in this study is sex. In their studies on obligation markers in different varieties of English, Tagliamonte and Smith (2006) and Tagliamonte and D'Arcy (2007) found differences between men and women in terms of variant choice, indicating that women tend to use newer forms. This supports patterns found in other sociolinguistic research such as Labov (1972). Based on this tendency, it is posited that the choice of variant among women in this group will reflect a more advanced phase in the grammaticalization process of these markers in comparison to the men. Specifically, they will be more likely to use *haber de* in epistemic contexts, and *tener que* in a wider variety of contexts in comparison to their male counterparts.

---

<sup>13</sup> An older and possibly separate group -those older than 70 years- rarely congregated in the streets like the other groups. Instead, they would remain in their homes and accept visitors. But, again, given the limited number of total participants, they were collapsed into the older adulthood group.

## 2.8 Summary

Taking into consideration that variation is structured and that these patterns can be attributed to certain linguistic and social factors, both a synchronic and a diachronic analysis of this modality system allows for explanations of patterns of modality in modern Spanish such as those observed in Bybee and Fleischman (1985), Marrano (1997), and Cornillie (2007). Furthermore, these results can also corroborate such typologies postulated by Bybee et al. (1994) and Olbertz (1998) or explicate discrepancies among the functions outlined for these modals.

Based on prior research, we can see that the English modal system is in flux and the research that is available on Spanish suggests a similar situation. A comparative synchronic and diachronic analysis can confirm suspicions of a similar process occurring in Spanish. Similarly, results gathered can support or refute claims of diachronic universality. Considering the dearth of research that has been conducted on Spanish, this study adds to the existing research on Spanish modals of obligation, none of which examine variation within the system. Furthermore, it will add to our understanding of both the history and the modern-day reality of Spanish grammar.

## CHAPTER 3 COMPETITION, VARIATION AND LANGUAGE CHANGE

### **3.1 Grammaticalization of Changing Forms**

Up until this point this study has demonstrated how the aforementioned constructions have developed into obligation markers in similar ways: they all originated as verbs that expressed possession and, as their frequency of use increased, they developed into obligation markers. Research that has focused on these transformations has often categorized these changes as grammaticalization (Meillet, 1912; Hopper & Traugott, 2003, 2). In addition, investigation into the processes involved in this change has shown that these forms undergo a general semantic-pragmatic shift as they move towards a more speaker-oriented and thus subjective meaning, in a process known as subjectification (Traugott, 1989; Bybee et al., 1994; Tagliamonte, 2004; Narrog, 2005; Tagliamonte & Smith, 2006; Tagliamonte & D'Arcy, 2007).

As these forms begin to undergo grammaticalization and move towards more subjective meanings, processes such as decategorialization, phonetic reduction, reanalysis, semantic bleaching or emancipation from the source form are likely to occur. Taking into consideration that grammaticalization is a gradual process, it is not required that a construction suffer each of these changes to be considered as having undergone grammaticalization. In fact, based on the inherent gradualness of grammaticalization, it is rare to have historical evidence showing a form going through each of these processes.

What is crucial to the present paper, however, is to identify whether any of the four aforementioned obligation markers has undergone some of these processes. Given that linguists such as Bybee et al. (1991), Hopper and Traugott (1993) and Bybee et al.

(1994) claim that the paths of grammaticalization of these types of modals are universal, it is also essential to identify whether these obligation markers follow this path and corroborate their claims. They posit that constructions denoting a notion of possession develop into obligation markers and eventually acquire an epistemic meaning. In Spanish, we already see evidence of this path within the obligation system.

The four constructions *haber que*, *tener que*, *haber de* and *tener de* all derive from a source that historically conveyed a notion of possession.<sup>1</sup> As was discussed in section 1.4.1, there is evidence of an emancipation of the constructions *haber que* and *haber de* from their source form, also indicative of grammaticalization. Their decategorialization from transitive verbs to obligation markers also point to both grammaticalization and subjectification. We also see with the advent of these markers the fusion with independent morphemes (*de/que*) to form collocations, which also indicates grammaticalization (Torres Cacoullos, 2006). Finally, a semantic pragmatic shift in each of these collocations from contexts where obligation is expressed to contexts where there is a more subjective meaning would imply grammaticalization.

Recall that in Chapter 1, after a cursory examination of these four collocations, it was postulated that these forms have undergone both subjectification and grammaticalization and, with the exception of *tener de*, continue to develop and change. This preliminary assessment suggested that Spanish modality develops like modality in other languages thus indicating that the changes found in the modal system in Spanish should follow the universal path postulated by Bybee et al. (1994).

---

<sup>1</sup> For more discussion on the development of these forms see Chapter 1, section 1.4.

If this is the case, the data should show the generalization of these constructions into new contexts, which would be evident if we find certain patterns of variation, possible decategorialization, or replacement of older constructions by newer ones. Additionally, an increased fusion of the independent morphemes of the grammaticalizing constructions should also be seen in the data through a decreased amount of examples with elements such as clitics or negation markers occurring between the components of the collocation. If an increase in frequency is observed along with these possible patterns of change (for a more detailed description on the mechanisms associated with grammaticalization see section 1.3.1.1), it would verify that these forms are evolving through grammaticalization.

Little work has been done on the grammaticalization of these expressions of obligation in Spanish. Taking each of these apparent changes into account, we examine these constructions and determine whether these forms are undergoing grammaticalization. In addition, we look at both the relative frequencies and the multivariate analyses of each of these constructions as possible corroboration that these forms are undergoing further grammaticalization. These results should depict the path that each of these markers takes as they expand into new contexts as obligation markers and eventually start to appear in epistemic contexts. This chapter describes the development of each of these markers by analyzing the tendencies of each construction over time.

One of the ways in which these tendencies are identified is through frequency. There is a large amount of research that cites the importance of frequency in language change (e.g. Haiman, 1994; Bybee, 2003a; Torres Cacoullos, 2006). In fact, Bybee

(2003a: 603) defines grammaticalization in relation to frequency by identifying it as ‘the process by which a frequently used sequence of words or morphemes becomes automated as a single processing unit.’

### 3.2 Frequency and Language Change

There are two types of frequency to consider when looking at a grammaticalizing construction: token and type frequency. Token, or text, frequency is the overall frequency of a construction in the data. Type frequency, on the other hand, refers to the variety of lexical elements that are used with a given construction. As was summarized in Table 2-1 in section 2.1.1 (repeated here in Table 3-1 for convenience) the token frequencies of these collocations reveal several patterns. For example, *tener que* occurs .11 times for every 5,000 words in the 16<sup>th</sup> century. By the 20<sup>th</sup> century, this number jumps to 2.5 times for every 5,000 words, more than 22 times more frequent in this century.

We also see a considerable increase in *haber que* from the 16<sup>th</sup> century to the 20<sup>th</sup> where, in the earlier century, it occurred .07 times for every 5000 words but increased to .65 in the 20<sup>th</sup>.<sup>2</sup> Although this is not as sharp of an increase as *tener que*, the number increased nine-fold which, again, is substantial. The opposite occurs with *haber de* in the 16<sup>th</sup> century where it occurs 3.58 times per 5000 words, but decreases to .99 in the 20<sup>th</sup> century.

Although some of these differences may be attributed to dialect, writing style and/or other social differences, these numbers are noticeably large enough to suggest

---

<sup>2</sup> Compare this to Luna Traill (1980) who, in her study on the use of periphrastic verbs in oral educated speech of Mexico City found that, of the obligation markers, *haber de* occurred in .4% of the data, *haber que* in 18.86% of the data and TQ in 51.01% (she also included *deber (de)*, which accounted for 29.71% of the data). Additionally, she did not find any examples of *tener que* being used as an epistemic marker (p 195).

that some change has taken place. Based on this change, we can assume that there are also changes in the type frequency, since an increase in token frequency can often be the result of an increase in type frequency (Bybee, 2003a).

Table 3-1. Absolute and relative frequency of each variant per century in the written corpora normalized per 5000 words

Century (word count)	<i>TQ</i> N	Normalized per 5,000	<i>TD</i> N	Norm	<i>HD</i> N	Norm	<i>HQ</i> N	Norm	Total
16 <sup>th</sup> (~2,339,244)	32	.11	89	.29	1,091	3.58	21	.07	1,233
17 <sup>th</sup> (~960,832)	40	.17	22	.10	852	3.70	19	.08	933
18 <sup>th</sup> (~403,340)	39	.50	12	.15	256	3.28	9	.12	316
19 <sup>th</sup> (~789,217)	253	1.00	2	.01	687	.67	79	.32	1,021
20 <sup>th</sup> (~3,501,511)	865	2.41	3	.01	354	.99	232	.65	1,454
Total N	1,563		128		3,248		381		4,957

To measure the changes in type frequency it is necessary to look at the distribution patterns of these constructions within each factor group. As these markers grammaticalize, they will start to generalize and occur with more lexical items (Bybee, 2003a). For example, in a study on deontic and epistemic modality in German, Heine (1995: 26) found that deontic readings were more associated with action verbs, interrogative sentences, first and second-person subjects, and past or present perfect tense. Given the small inventory of lexical items with which these deontic modals usually occur, their type frequency is relatively low. But as these modals start to increase in frequency they will start to generalize and be used in more contexts, namely those associated with epistemic readings.

A similar trend is expected in the current study. In the earlier centuries the newer forms should have relatively low type frequencies but will generalize into new contexts as time progresses. A detailed look at each of these factor groups across centuries should uncover these gradual changes. In each subsection below a table is provided to look at the frequency patterns of each variant across time within a particular factor

group. Each century has two columns; one is labeled 'N' or number. This is the relative frequency of occurrence of the variant in this particular context in comparison with the other three variants. The other column labeled 'percentage' represents the percentage of occurrences of the variant in the given context compared to the other variants. Each variant is discussed separately in relation to the frequency of occurrence within each factor group.

### **3.2.1 Verb Type**

As we saw in Chapter 2, verb type can have an effect on the interpretation of a modal verb. For example, action verbs are more likely to be found in contexts with deontic modals than with epistemic ones (Heine, 1995, Silva-Corvalán, 1995; Cornillie, 2007). In addition, older more grammaticalized constructions (e.g. *haber de*) are more likely to occur with stative verbs. This is confirmed by such studies as Coates (1983), Pietrandrea (2005); and Cornillie (2007). The patterns of frequency within this group type can be an additional tool to help illuminate the history of these constructions. In the following sections we partially reconstruct the patterns of use of each of these obligation markers.

#### **3.2.1.1 *Tener que***

Table 3-5 reveals that as time passes *tener que* starts appearing with a wider variety of verb types. In the earlier centuries *tener que* is relatively limited in terms of contexts. Although the highest percentages of occurrence are found with perception verbs (7.1%) and communication verbs (17.3%), there is also a high number of examples with action verbs (n=13). As was suggested earlier, this is an area in the grammar associated with deonticity and may be an entry point for the then new construction. In fact, in terms of raw percentage *tener que* occurs the most often with

action verbs in every century, despite this not being reflected in the overall relative frequency percentages. One area that has relatively low overall frequency but is increasingly used with *tener que* is perception verbs. Cornillie (2007) found that when *tener que* appeared with a perception verb it was always in a deontic context. This association may be due to the historical association of this construction with perception verbs.

In the earlier centuries *tener que* does not appear in contexts with copular verbs and only minimally in contexts with psychological verbs. Around the 17<sup>th</sup> and 18<sup>th</sup> centuries, there is an increase in the frequency of *tener que* in contexts with psychological verbs (14%), but that increase is practically halted during the 19<sup>th</sup> century and up until the 20<sup>th</sup>, where we see a dramatic increase from 16% to 51.5%.

Table 3-2. Frequency of *tener que* with the factor group verb type

Verb Type	16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
	N	%	N	%	N	%	N	%
action	13	2.2	22	3.6	123	31.6	369	64.7
<i>hacer</i>	4	4	17	18.7	16	25.8	58	73.4
perception	1	7.1	11	42.3	7	33.3	62	82.7
communication	9	17.3	8	15.1	25	30.1	93	65.5
psychological	2	2.8	11	14.1	30	16.1	124	51.5
<i>haber/tener</i>	0	0	0	0	2	7.7	5	20
copular	0	0	1	.7	10	8.2	51	39.2
influence	0	0	0	0	2	33.3	7	87.5
other	3	2.2	9	4.1	38	30.4	96	52.2

A similar trend is seen with contexts containing copular verbs, although the initial increase is not seen until the 19<sup>th</sup> century and is minimal compared to that of psychological/emotion verbs. *Tener que* only accounts for 39% of the contexts containing copular verbs in the 20<sup>th</sup> century. This is relatively low compared to its occurrence with all other verbs types (with the exception of the verbs *haber* and *tener*). This result may imply that, as this construction grammaticalizes and moves towards a

more epistemic meaning, it will occur more and more often with copular verbs (Pietrandrea, 2005).

### 3.2.1.2 *Haber de*

The oldest construction in the group, *haber de*, is used almost exclusively during the 16<sup>th</sup> century when it comes to verb type (Table 3-3). It occurs in over 80% of the contexts, regardless of the type of verb. The only exception to this is with communication verbs. Based on the data, communication verbs may have been an entry point for the other obligation markers and thus were the source of the diminishing control *haber de* has over the system over the subsequent centuries. This is particularly evident with perception verbs. As *haber de* decreases in total percentage in this group *tener que* seems to take over these contexts. This may be the advent of the spread of *tener que* into new contexts. Aside from this, *haber de* is used decreasingly across time in general. In fact, excluding contexts with the verbs *haber* or *tener*, *haber de* is used less than 50% of the time, regardless of the verb type, during the 20<sup>th</sup> century. This is in stark contrast to the 16<sup>th</sup> century.

Another notable increase occurs in the period between the 16<sup>th</sup> century and the 17<sup>th</sup> and 18<sup>th</sup> centuries with copular verbs. By the 17<sup>th</sup> and 18<sup>th</sup> centuries *haber de* has almost monopolized contexts with copular verbs (99.3%). This diminishes slightly, however, in the 19<sup>th</sup> century as *tener que* starts to appear more often in these contexts. An even more dramatic change in occurrence with copulas is noted in the 20<sup>th</sup> century as *haber de* decreases from 87.7% to 44.6%. This decrease is particularly important because the literature identifies stative verbs as verb types that commonly occur in epistemic contexts (e.g. Bybee et al. 1994; Pietrandrea, 2005; Aaron, 2006; Cornillie, 2007;) and copular verbs are prototypically the most common of the stative verbs

(Pietrandrea, 2005). This implies that, by the 20<sup>th</sup> century, *tener que* is not only taking over the obligation system, but also starting to move into possible epistemic contexts that have historically been reserved for *haber de*. A similar situation has occurred in the English obligation system. In Toronto English the obligation marker *have to* has virtually taken over the entire obligation system and, in certain dialects of British English, *have to* is being used in contexts historically reserved for epistemic *must* (Tagliamonte & Smith 2006).

Table 3-3. Frequency of *haber de* with the factor group verb type

Verb Type	16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
	N	%	N	%	N	%	N	%
action	543	89.9	554	91.6	239	61.4	101	17.7
<i>hacer</i>	82	82.8	68	74.7	42	67.7	5	6.3
perception	12	85.7	14	53.8	10	47.6	9	12
communication	26	50	40	75.5	48	57.8	29	20.4
psychological	62	87.3	57	73.1	132	71.0	80	33.2
<i>haber/tener</i>	88	95.7	36	100	23	88.5	13	52
copular	146	96.1	134	99.3	107	87.7	58	44.6
influence	9	90	6	100	3	50	1	12.5
other	123	88.5	198	90.4	82	65.6	58	31.5

### 3.2.1.3 *Haber que*

Although *haber que* appears as an obligation marker quite early in Spanish (Yllera, 1980) it, like *tener que*, tends to appear primarily in contexts with communication verbs during the 16<sup>th</sup> century (21.2%), as is illustrated in Table 3-4. Again, this further supports the notion that communication verbs are an entry point for these newer obligation markers. Other than the examples with communication verbs, this marker is rarely used in the 16<sup>th</sup> century. But even the occurrences with communication verbs drop during the 17<sup>th</sup> and 18<sup>th</sup> centuries by 17%. However, as the use of *haber que* decreases in contexts with communication verbs, the construction spreads to new territory. This is evidenced by a 9% increase with psychological verbs and a 5%

increase with action verbs. Recall that an overall similar trend was found with *tener que*: communication verbs as the entry point, then a subsequent spread into new contexts.

During the 19<sup>th</sup> century *haber que* sees a small but evident increase of occurrence with all verb types. The extension isn't as dramatic as the increase found with *tener que* but it is still clear that this construction is being used in a wider variety of contexts. As it competes with the other obligation markers for territory in the 19<sup>th</sup> century, it also becomes increasingly used with verbs of perception. This is based on the 15% increase seen in contexts with *haber que* between the 17<sup>th</sup> and 18<sup>th</sup> centuries and the 19<sup>th</sup> century. This trend changes, however, in the 20<sup>th</sup> century as *tener que* begins to take over the entire system. *Haber que* sees a 15% decrease in occurrence with perception verbs during this period. Nevertheless, *haber que* continues to be used in the 20<sup>th</sup> century but its use spreads into other contexts. In fact, with the exception of verbs of influence and perception, there is an increase in the frequency of use of *haber que* with all types of verbs.

Table 3-4. Relative frequency of *haber que* with the factor group verb type

Verb Type	16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
	N	%	N	%	N	%	N	%
action	5	.8	11	1.8	26	6.7	100	17.5
<i>hacer</i>	2	2	4	4.4	4	6.5	16	20.3
perception	0	0	0	0	4	19	3	4
communication	11	21.2	2	3.8	10	12	20	14.1
psychological	1	1.4	8	10.3	24	12.9	37	15.4
<i>haber/tener</i>	0	0	0	0	1	3.8	7	28
copular	1	.7	0	0	4	3.3	20	15.4
influence	0	0	0	0	1	16.7	1	12.5
other	1	.7	3	1.4	5	4	29	15.8

### 3.2.1.4 *Tener de*

As already mentioned *tener de* essentially falls out of use in the 19<sup>th</sup> century (Table 3-5). This is evidenced by a steady decrease throughout the 16<sup>th</sup>, 17<sup>th</sup> and 18<sup>th</sup>

centuries and the relatively limited number of occurrences of this construction in every century. Early on, during the 16<sup>th</sup> century *tener de* does occur occasionally in contexts with all verb types but is used most often with communication verbs (11.5%) and the verb *hacer* (11.1%). This quickly changes during the 17<sup>th</sup> and 18<sup>th</sup> centuries, where there is a shift towards using *tener que* and *haber de* with the verb *hacer* and also a large increase in the use of *haber de* with communication verbs.

- 3.1 *como chapulín, de aquí para allá, aunque ya **tengo de** estarme quieto como quince años.* (Martín del Campo, 1976: CORDE)  
 ‘like a locust, from here to there, even though I already have to stay still like a 15 year old.’

At the turn of the 19<sup>th</sup> century *haber que* and *tener que* continue to compete with *haber de* but *tener de* becomes all but obsolete, occurring only in a few isolated cases (3.1).

This trend continues into the 20<sup>th</sup> century suggesting that this form is no longer part of the obligation system.

Table 3-5. Frequency of *tener de* with the factor group Verb type

Verb Type	16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
	N	%	N	%	N	%	N	%
action	43	7.1	18	3	1	.3	0	0
<i>hacer</i>	11	11.1	2	2.2	0	0	0	0
perception	1	7.1	1	3.8	0	0	1	1.3
communication	6	11.5	3	5.7	0	0	0	0
psychological	6	8.5	2	2.6	0	0	0	0
<i>haber/tener</i>	4	4.3	0	0	0	0	0	0
copular	5	3.3	0	0	1	.8	1	.8
influence	1	10	0	0	0	0	0	0
other	12	8.6	9	4.1	0	0	1	.5

### 3.2.2 Grammatical Person and Animacy

Grammatical person can be a strong indicator of the grammaticalization of these constructions. In separate studies of the English modal system Coates (1983) and Pietrandrea (2005) find that future epistemics are more likely to occur with stative verbs.

Bybee and Fleischman (1995) find a similar pattern with modals and claim that they are more likely to be agent-oriented (i.e. deontic) when the subject is in first-person and more likely to be epistemic when it is in third-person. This would imply that, as these forms grammaticalize and move towards epistemicity, they will more likely start to appear in contexts where the subject is in third-person.

Animacy also plays a role in this process. Recall that in his study on the evidential and epistemic readings of various Spanish semi-auxiliaries Cornillie (2007:219) found that *tener que* occurs as a deontic construction (97.75%) if an animate subject is present. Coates (1983) and Heine (1995) also suggest that animacy is a determining factor in whether a reading is epistemic or deontic. They suggest that inanimate subjects are more associated with epistemicity than animate subjects. Wårnsby (2004: 179) echoes this claim but extends it by adding that existential subjects are also associated with epistemicity.

This seems to be partially true in Spanish where, regardless of the obligation marker used, an epistemic meaning can be gleaned, as in (3.2), but it is not an exclusively epistemic context. Nevertheless, it does appear to be a context where either meaning can be extracted and thus a possible area in the grammar for identifying the gradual shift into epistemic contexts. Thus, we can assume that as these forms grammaticalize over time, they will slowly begin to shift away from contexts in first and second-person and move towards contexts with inanimate or existential third-person subjects.

- 3.2 *porque de aquí adelante no ha de haber abdiencia en Panamá. Otrosí: mandamos que \*\** (Anónimo, 1544; CORDE)  
'because from today going forward there must not be an audience in Panamá'

### 3.2.2.1 *Haber de*

Following a pattern similar to those of the earlier discussed factor groups, we can see in Table 3-6 that, during the 16<sup>th</sup> century, *haber de* is the dominant form in every grammatical person except first (singular and plural) (36%). This could be described as the entry point for new forms in terms of grammatical person. *Haber de* is, however, increasingly used in first-person through the 17<sup>th</sup>-19<sup>th</sup> centuries but this tendency is halted in the 20<sup>th</sup> century where we see a decline of 40.9% points when compared to the occurrences during the 19<sup>th</sup> century.

Knowing that epistemicity is often associated with third-person inanimate subjects, it is surprising that *haber de* continues to be found less and less frequently in these contexts over the centuries, accounting for only 47.9% of cases in the 20<sup>th</sup> century. This could be an indication that, because this construction is moving into primarily epistemic contexts, as Cornillie (2007) suggests, *tener que* is expanding into its territory, even the territory is associated with epistemic meanings. This could be the beginning of the end of *haber de* as an obligation marker.

Table 3-6. Frequency of *haber de* with the factor group grammatical person

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Grammatical Person	1 <sup>st</sup>	31	36	102	72.9	195	69.9	60	19
	2 <sup>nd</sup>	110	97.3	62	91.2	72	84.7	51	38.6
	3 <sup>rd</sup> animate	658	89.8	620	87.7	284	57.1	157	19
	3 <sup>rd</sup> inanimate	240	97.6	284	96.6	130	86.7	80	47.9
	animated objects	20	90.9	32	97	3	42.9	1	20
	existential	32	97	7	100	2	100	5	71.4

Even though Wårnsby (2004) is focusing on English when he determines that existential subjects are associated with epistemicity, we can see a somewhat similar

trend in Spanish. In the earlier centuries, existential subjects are associated almost exclusively with *haber de*. This trend continues through the 19<sup>th</sup> century. Even in the 20<sup>th</sup> century, the majority of contexts with existential subjects are still used with *haber de* (71.4%), although there is a decline, indicating that this may be one of the last areas into which *tener que* will expand.

### 3.2.2.2 *Tener de*

In the 15<sup>th</sup> and 16<sup>th</sup> centuries it is evident that *tener de* was principally used in contexts with first-person subjects (Table 3-7). Consequently, we can conclude that this form was more connected with an obligation reading than an epistemic one. This is expected given the relative youth of this form in comparison to *haber de*. The strong link to obligation is further supported by the minimal amount of occurrences of this marker with other grammatical persons. Nevertheless, *tener de* is found in contexts with all grammatical persons with animate or inanimate subjects, which suggests that this form was perhaps grammaticalizing and moving towards a more subjective meaning before it

Table 3-7. Frequency of *tener de* with the factor group grammatical person

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Grammatical Person	1 <sup>st</sup>	43	50	22	15.7	1	.4	1	.3
	2 <sup>nd</sup>	2	1.8	0	0	0	0	1	.8
	3 <sup>rd</sup> animate	35	4.8	11	1.6	1	.2	1	.1
	3 <sup>rd</sup> inanimate	6	2.4	2	.7	0	0	0	0
	animated objects	2	9.1	0	0	0	0	0	0
	existential	1	3	0	0	0	0	0	0

fell out of use. The marker appears to stop grammaticalizing, however, in the 17<sup>th</sup> and 18<sup>th</sup> centuries where it occurs primarily in contexts with a first-person subject. Those remaining examples that don't occur with first-person subjects are predominantly associated with third-person animate subjects, which is not associated with epistemicity

to the same degree as third-person inanimate subjects. By the 19<sup>th</sup> century even its use with first-person subjects dwindles to a mere 4% of all contexts with first-person subjects. The diminishing frequency and use of this form is likely owed to the ever increasing use of *tener que*.

### 3.2.2.3 *Tener que*

As a new obligation marker, we would expect *tener que* to be used primarily in contexts with first or second-person, given that these are generally associated with obligation. In the 16<sup>th</sup> century we see that *tener que* accounts for 14% of all contexts with first-person subjects, and only 2.7% of both third-person animate and inanimate subjects (Table 3-8). This is an expected result. What this suggests is that this newer construction is entering the obligation system via contexts that the older forms have moved out of. This would imply, though, that *haber de* has already started to grammaticalize and move towards areas in the grammar that are associated with epistemicity in the 16th century (e.g. contexts with third-person inanimate subjects).

In the 17<sup>th</sup> and 18<sup>th</sup> centuries there is an increase in the amount of times *tener que* occurs in third-person animate subject contexts, which continues into the later centuries. We also see a decreased use of *tener que* with first-person subjects during the 17<sup>th</sup> and 18<sup>th</sup> centuries, although it's minimal. By the 20<sup>th</sup> century this marker is the predominant construction used in contexts with first-person subjects, accounting for 80.1% of all cases.

What is equally apparent during the 17<sup>th</sup> and 18<sup>th</sup> centuries is the increase of the use of *tener que* in contexts with second-person subjects. Given that Coates (1983) suggests that second-person subjects encode the strongest obligation reading, we can assume that *tener que* takes on a stronger obligation meaning early on. This continues

all the way through the 20<sup>th</sup> century where *tener que* accounts for 59.8% of all contexts with second-person subjects.

Table 3-8. Frequency of *tener que* with the factor group grammatical person

Grammatical Person		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Person	1 <sup>st</sup>	12	14	16	11.4	81	29	253	80.1
	2 <sup>nd</sup>	0	0	6	8.8	13	15.3	79	59.8
	3 <sup>rd</sup> animate	20	2.7	48	6.8	137	27.6	442	53.4
	3 <sup>rd</sup> inanimate	20	2.7	8	2.7	18	12	85	50.9
	animated objects	0	0	1	3	4	57.1	4	80
	existential	0	0	0	0	0	0	2	28.6

All these tendencies imply that *tener que* has been established as an obligation marker and has essentially taken over the territory of both *haber de* and *tener de* by the 20<sup>th</sup> century in terms of grammatical person and animacy. We can also see that *tener que* is grammaticalizing and moving into epistemic contexts by the steady increase of cases of *tener que* occurring with third-person subjects. By the 20<sup>th</sup> century it accounts for over half of the cases with both animate and inanimate subjects. Although this suggests that this marker has grammaticalized and is becoming more associated with epistemicity, it is still used minimally with the epistemic-related existential subjects, where it accounts for 28% of all cases. This area in the grammar is still primarily associated with *haber de*.

### 3.2.2.4 *Haber que*

In terms of grammatical person, there are very limited examples where *haber que* (Table 3-9) is used with anything but non-referential (i.e. third-person animate) subjects. Of course, this is due to the fact that it can only occur with animate subjects and it is always conjugated in third-person. Nevertheless, there are several isolated examples of this obligation marker being use with subjects other than third-person (3. 3).

- 3.3 *Los podremos ayudar para que puedan hacer alguna carrera. Total, no hay que precipitarnos...*(Azuela, 1985: CREA)  
 ‘We will be able to help them so that they can run some race. Anyway, we don’t have to rush.’

That being said, research indicates that grammatical person and animacy are some of the grammatical components that measure or can be used to determine the age of deontic modals (e.g. Cornillie, 2007). Based on the limitations of this construction, it may be that this form can not move into epistemic contexts and thus is limited to deonticity, as suggested by Cornillie, De Mulder and Vermandere (2009: 115): ‘unlike *haber de*, the impersonal construction *hay que* is restricted to a deontic reading.’ If this is the case, then this should be reflected in the results.

Table 3-9. Frequency *haber que* with the factor group grammatical person

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Grammatical Person	1 <sup>st</sup>	0	0	0	0	2	.7	2	.6
	2 <sup>nd</sup>	1	.9	0	0	0	0	1	.8
	3 <sup>rd</sup> animate	20	2.7	28	4	77	15.4	229	27.6
	3 <sup>rd</sup> inanimate	0	0	0	0	0	0	0	0
	animated objects	0	0	0	0	0	0	0	0
	existential	0	0	0	0	0	0	0	0

### 3.2.3 Sentence Type

Coates (1983) and Bybee and Fleischman (1995) illustrate that in English there are limitations to epistemic modals occurring in interrogative sentences. Due to the limited amount of research done on the grammaticalization of obligation markers in Spanish, there is little evidence that a similar trend occurs in Spanish. A detailed look at each of these constructions may answer this question.

#### 3.2.3.1 *Haber de*

The oldest of the four constructions, *haber de*, shows tendencies to be the marker of choice in interrogative sentences up until the 19<sup>th</sup> century (Table 3-10). Interrogative

sentences, which only account for 6% of the data in the 16<sup>th</sup> century, 10% of the data in the 17<sup>th</sup> and 18<sup>th</sup> centuries, and 18% of the data in the 19<sup>th</sup> century occur with *haber de* in over 80% of the cases. There is an overall decline in the number of interrogative sentences, however, in the 20<sup>th</sup> century, where it accounts for only 6% of the data. During this century *haber de* only occurs with interrogative sentences 40% of the time, while *tener que* occurs 53% of the time. Thus, we can conclude that this decrease in *haber de* in interrogative sentences is likely due to loss of territory to *tener que*. It may also indicate the movement of *haber de* into more epistemic contexts however it's hard to conclude that just based on this data.

Table 3-10. Frequency *haber de* with the factor group type of sentence

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Sentence Type	declarative	1028	88.9	997	88.2	509	61.3	318	23.3
	interrogative	63	82.9	110	92.4	177	93.7	36	40

### 3.2.3.2 *Tener de*

As can be seen in Table (3-11), in the 16th century *tener de* accounted for a larger percentage of interrogative sentences (17.1%) than declarative ones (6.6%). This may suggest that because *haber de* was grammaticalizing and thus being used in epistemic contexts, *tener de* was beginning to be the marker of choice in interrogative sentences,

Table 3-11. Frequency *tener de* with the factor group type of sentence

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Sentence Type	declarative	76	6.6	31	2.7	2	.2	2	.1
	interrogative	13	17.1	4	3.4	0	0	1	1.1

since epistemicity would imply that *haber de* would be used in limited ways in these types of contexts (Coates, 1983). This is hard to substantiate, however, since *tener de*

falls out of use before it has the opportunity to completely replace *haber de* in these contexts.

### 3.2.3.3 *Tener que*

This marker was rarely used in interrogative sentences until the 20<sup>th</sup> century, at which point there is a large increase in the number of cases of *tener que* in these contexts (Table 3-12). This further suggests that interrogative sentences may not be indicators of grammaticalization in Spanish, since we would expect *tener que* to occur in the interrogative sentences early on and gradually, as it undergoes subjectification, this would start to diminish.

Table 3-12. Frequency *tener que* with the factor group type of sentence

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Sentence Type	declarative	32	2.8	74	6.5	245	29.5	817	59.9
	interrogative	0	0	5	4.2	8	4.2	48	53.3

### 3.2.3.4 *Haber que*

Similar to *tener que*, *haber que* is seldom used in interrogative sentences from the earlier centuries through the 20<sup>th</sup> century (Table 3-13). Again, this could indicate grammaticalization, but this is difficult to determine given the relatively low number of interrogative sentences in general (n =90) and the even lower number of *haber que* occurring in interrogative contexts (n=5).

Table 3-13. Frequency *haber que* with the factor group type of sentence

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Sentence Type	declarative	21	1.8	28	2.5	75	9	227	16.6
	interrogative	0	0	0	0	4	2.1	5	5.6

### 3.2.4 Polarity

Recall that in section 2.6.4 that I discussed how negative polarity is less associated with epistemicity than with deonticity due to the limitations of negation in epistemic contexts (Coates, 1983; Palmer, 1990; Heine, 1995; Silva-Corvalán, 1995). Therefore, the inclination of a construction to not occur with negative polarity is likely attributed to the movement of this construction into epistemic contexts.

As can be seen in Table 3-14, *haber de* shows a tendency to be used increasingly less with negation as time passes. This could reflect its movement into epistemic contexts and thus echo the claims made by the aforementioned researchers. We can assume, then, that contexts of negation may be a good entry point for new obligation markers. This appears to be the case with *tener que* and *haber que*. They both appear to continually increase in frequency of occurrence in contexts with negative polarity. However, during the 19<sup>th</sup> century *tener que* decreases by about 6%, while *haber que* increases by about the same amount. It appears, then, that at this point in time *haber que* seems to be taking over these contexts at a faster rate. By the 20<sup>th</sup> century the

Table 3-14. Frequency of all constructions with the factor group polarity

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
<i>haber de</i>	positive	987	89.8	993	92.2	551	67.3	309	23.6
	negative	104	77.6	114	66.3	135	67.2	45	30.8
<i>tener de</i>	positive	80	7.3	34	3.2	2	.2	3	.2
	negative	9	6.7	1	.6	0	0	0	0
<i>tener que</i>	positive	21	1.9	41	3.8	221	27	799	61.1
	negative	11	8.2	38	22.1	32	15.9	66	45.2
<i>haber que</i>	positive	11	1	9	.8	45	5.5	197	15.1
	negative	10	7.5	19	11	34	16.9	35	24

increase of *haber que* has been overshadowed by the dramatic increase of *tener que* in contexts with negation. There is also a sharp decline in frequency of *haber de* with negative polarity in the 20<sup>th</sup> century. However, while *tener que* accounts for 45.2% of

contexts with negation, there is still plenty of territory associated with the other two markers. Taking this into consideration, it may be that negation is not a conditioning factor when determining which of these markers will be used. This is discussed in more detail in Chapter 4.

### 3.2.5 Preceding Clitic

In section 2.3.6 we discussed Myhill's (1998) study, which indicated that clitic climbing occurs more often with auxiliary verbs rather than lexical verbs. Although the constructions in the present study may not have reached auxiliarihood yet (Olbertz, 1998), they are often referred to in the literature as both modal and semi-auxiliaries. As these forms grammaticalize and move toward auxiliarihood there should be evidence of more clitic climbing. We would expect to see an increase of occurrences with *haber de*, or in the later centuries with *tener que*, in contexts with a preceding clitic. If we look at Table 3-15, we see that there is a decrease in the total percentage of preceding clitics found with *haber de*. When we look at *tener que* we see that it increasingly occurs with preceding clitics as time goes by, although the difference in actual numbers between preceding clitic and others is staggering (79 vs.786 respectively).

Table 3-15. Frequency of all constructions with the factor group preceding clitic

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
<i>haber de</i>	yes	406	94.2	391	97.5	207	99	67	45.9
	other	685	85.4	716	84.4	479	59.1	287	21.9
<i>tener de</i>	yes	24	5.6	9	2.2	0	0	0	0
	other	65	8.1	26	3.1	2	.2	3	.2
<i>tener que</i>	yes	1	.2	1	.2	2	1.0	79	54.1
	other	31	3.9	78	9.2	251	30.9	786	60.1
<i>haber que</i>	yes	0	0	0	0	0	0	0	0
	other	21	2.6	28	3.3	79	9.7	232	17.7

These results suggest that these obligation markers are still not fully grammaticalized and thus are not processed as auxiliaries yet. This is supported by

Bybee et al. (1994:241) who suggest that agent-oriented modality (i.e. obligation) is at the beginning stages of grammaticalization. This is the other likely reason why preceding clitics do not comprise the majority of the cases.

### **3.2.6 Direct Objects**

Bybee et al. (1994) and Diaz-Campos (2011) echo similar claims as Klein Andreu (1991) in suggesting that one of the ways to identify semantic changes, particularly with constructions expressing modality, is by looking at low- or high-focus clauses. In her study on the grammaticalization of the pluperfect indicative in Latin and the imperfect subjunctive in Spanish, Klein Andreu (1991) identifies low-focus clauses as being associated with the older form while high-focus clauses was associated with the newer form. She defines high-focus clauses as those that present new information and occur with animate subjects or objects. She goes on to identify low-focus clauses as those that are identified through the presence of stative verbs and negation and often occurring in relative clauses.

If we apply the results of Klein Andreu (1991) to the case of the direct object we could posit that if it is animate the construction may be less grammaticalized since it is more likely to occur in high-focus clauses. In the 15-18<sup>th</sup> centuries, then, *haber de* is expected to occur with inanimate objects more often, whereas the new constructions are more likely to be used with animate objects. As we move towards the 20<sup>th</sup> century there should be a shift in *tener que* as it becomes more and more frequent, possibly indicating grammaticalization and therefore may be more likely to be found with inanimate objects.

### 3.2.6.1 *Haber de*

As previously mentioned, we would expect *haber de* to occur more often with inanimate objects as time progresses, since this is considered characteristic of a construction in a low-focus clause. In the 16<sup>th</sup> century we see that *haber de* occurs more often with inanimate objects (n=247) than with any other type of objects. This trend continues until the 20<sup>th</sup> century, at which point the construction *tener que* experiences a dramatic increase in frequency (see Table 3-16) and *haber de* suffers a parallel drop. Nevertheless, there are still more overall cases of *haber de* with inanimate objects in the 20<sup>th</sup> century than with any other variant, although *tener que* has only slightly less (92 vs. 88 respectively). Despite this, the percentages show a decrease in the frequency of use of *haber de* with inanimate objects when compared to the other variants, particularly *tener que*. This could be an early indication of *tener que* taking over the system and its movement from high-focus clauses to low-focus clauses.

Table 3-16. Frequency of *haber de* with the factor group direct objects

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Direct Objects	oblique	122	87.1	157	84.9	51	45.5	50	17.2
	copular	151	96.2	137	99.3	109	87.2	55	45.5
	inanimate	247	89.5	322	85.4	197	63.5	92	17.6
	animate	78	85.7	104	93.7	97	71.9	42	27.3
	intransitive <i>se</i>	40	88.9	35	92.1	34	87.2	6	46.2
	middle voice	227	96.2	172	96.6	90	69.2	23	46
	other	226	78.5	180	81.1	108	63.9	86	28.4

There is also an increase in the percentage of occurrences in which *haber de* is found with copular complements. This increase continues until the 19<sup>th</sup> century where, again, we start to see a decline. Similar increases are seen during the same period with complements expressed through intransitive *se* (e.g. *irse*), as well as with animate

objects but, like the copular complements, the occurrences begin to decline by the 19<sup>th</sup> century.

### 3.2.6.2 *Tener de*

During the 16<sup>th</sup> century, *tener de* appears to be used in all contexts (Table 3-17). In comparison to the other markers, it occurs more often in contexts with animate objects, inanimate objects, contexts with intransitive *se* (e.g. *irse*) or in contexts where a direct object isn't present. This tendency starts to change in the 17<sup>th</sup> and 18<sup>th</sup> centuries where there is a slight shift towards more occurrences in which an oblique complement is present. However, it persists in its occurrence with intransitive *se*. Despite the shift towards oblique complements, the percentage of the cases where we find *tener de* occurring with oblique complements is lower than it was in the previous two centuries.

Table 3-17. Frequency of *tener de* with the factor group direct objects

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Direct Objects	oblique	7	5	7	3.8	0	0	1	.3
	copular	5	3.2	0	0	1	.8	1	.8
	inanimate	23	8.3	19	5	1	.3	1	.2
	animate	11	12.1	3	2.7	0	0	0	0
	intransitive <i>se</i>	4	8.9	2	5.3	0	0	0	0
	middle voice	9	3.8	1	.6	0	0	0	0
	other	30	10.4	3	1.4	0	0	0	0

This may not be categorized as a shift towards oblique complements, but rather one of the last strongholds of this form as it slowly starts to fall out of use. By the 19<sup>th</sup> century *tener de* only occurs two times in all the data, thus further illustrating its demise.

### 3.2.6.3 *Tener que*

As *tener que* starts to be used as an obligation marker during the 16<sup>th</sup> century, it first starts to appear in contexts without direct objects or with oblique complements (Table 3-18). There are, however, a few isolated cases where it occurs with intransitive

se (n=1) or inanimate objects (n=3). This begins to change in the 17<sup>th</sup> and 18<sup>th</sup> centuries where the percentage of cases of *tener que* occurring with inanimate objects increases from 1.1% to 8.2%. Surprisingly, we don't see many contexts where this construction is

Table 3-18. Frequency of *tener que* with the factor group direct objects

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Direct Objects	oblique	8	5.7	15	8.1	50	44.6	200	69
	copular	0	0	1	.7	11	8.8	51	42.1
	inanimate	3	1.1	31	8.2	92	29.7	323	61.8
	animate	0	0	1	.9	30	22.2	88	57.1
	intransitive se	1	2.2	1	2.6	4	10.3	7	53.8
	middle voice	0	0	4	2.2	26	20	27	54
	other	20	6.9	26	11.7	40	23.7	169	55.8

used with animate objects until the 19<sup>th</sup> century. This suggests that, although animate objects may be associated with high-focus clauses and thus newer forms, it may not be an entry point for new obligation markers. Instead, these newer forms may first occur in contexts where no object is required and then, once it's established it moves into contexts with objects. By the 20<sup>th</sup> century, *tener que* accounts for over half of all cases with or without an object with the exception of copular objects. This might indicate that contexts with copulas are one of the last areas in the grammar for obligation markers to acquire.

#### 3.2.6.4 *Haber que*

*Haber que* and *tener que* appear to have similar use patterns during the 16<sup>th</sup> century, although *haber que* is more often used in cases with oblique complements or contexts where a direct object is not present (Table 3-19). One difference between the two markers is that *haber que* is used with animate objects in the earlier centuries. This could indicate *haber que*'s age since this construction appears as an obligation marker earlier than *tener que*. Or it may be that, because *haber que* had already come to be

associated with these contexts, *tener que* entered the obligation system being used in other contexts.

Table 3-19. Frequency of *haber que* with the factor group direct objects

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Direct Objects	oblique	3	2.1	6	3.2	11	9.8	39	13.4
	copular	1	.6	0	0	4	3.2	14	11.6
	inanimate	3	1.1	5	1.3	20	6.5	107	20.5
	animate	2	2.2	3	2.7	8	5.9	24	15.6
	intransitive <i>se</i>	0	0	0	0	1	2.6	0	0
	middle voice	0	0	1	.6	14	10.8	0	0
	other	12	4.2	13	5.9	21	12.4	48	15.8

During the 17<sup>th</sup> and 18<sup>th</sup> centuries there is little change in terms of *haber que* occurring with new types of objects, although there is a slight increase in the percentage of cases of *haber que* in the contexts in which it was already occurring in previous centuries. In the 19<sup>th</sup> century this increase continues and we see an expansion of *haber que* into new contexts, like objects expressing middle voice (10.8%), inanimates (6.5%), oblique complements (9.8%) and, when a direct object isn't present (12.4%). There continues to be an increase in the use of this marker with all types of objects with the exception of intransitive *se* and middle voice. Nevertheless, the percentage of occurrences of *haber que* in each of these cases is minimal in comparison to those of *tener que*.

### 3.2.7 Indirect Objects

Similar to direct objects, this factor group was primarily chosen to determine the age of the obligation markers and where in the grammaticalization process they are. I postulated that indirect objects can help identify whether these markers are in high- or low-focus clauses through animacy. In each of the data sets every example of an indirect object referred to an animate being. As a result, each occurrence was marked

as either having an indirect object or not. Based on Klein Andreu's (1991) analysis of subjunctive forms, animate objects should be associated with the newer forms.

Therefore, in the earlier years we expect to see *haber de* more often in contexts with indirect objects but there should be a decreasing number of cases in which it appears with indirect objects as time progresses. In addition, we would expect the presence of an indirect object to be a possible entry point for new forms.

Just as expected, in the 16<sup>th</sup> century *haber de* accounts for 87.8% of all contexts with indirect objects. *Tener de* is found in 8.5% of the data, while *tener que* only accounts for 3.7% and *haber que* doesn't occur with indirect objects (Table 3-20). *Haber de* maintains its hold during the 17<sup>th</sup> and 18<sup>th</sup> centuries but starts to be used less (71.8%) in the 19<sup>th</sup> century as *tener que* and *haber que* start to be employed in its place. By the 20<sup>th</sup> century *haber de* only accounts for 2.4% of the cases with indirect objects as *tener que* becomes the dominant choice (57.8%) and *haber que* is found in 17.6% of

Table 3-20. Frequency of all factors occurring with the factor group indirect objects

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
<i>haber de</i>	present	72	87.8	101	87.8	56	71.8	84	24.3
	not present	1019	88.5	1006	88.7	630	66.9	270	24.4
<i>tener de</i>	present	7	8.5	9	7.8	0	0	1	.3
	not present	82	7.1	26	2.3	2	.2	2	.2
<i>tener que</i>	present	3	3.7	4	3.5	19	24.4	200	57.8
	not present	29	2.5	75	6.6	234	24.8	665	60
<i>haber que</i>	present	0	0	1	.9	3	3.8	61	17.6
	not present	21	1.8	27	2.4	76	8.1	171	15.4

cases. These results illustrate that *haber de* has lost its territory to the grammaticalizing *tener que*. This could indicate that *haber de* has grammaticalized and is showing evidence of appearing in low-focus clauses in the later centuries. This claim is slightly diminished when an equally low percentage is found with *haber de* when an indirect object is not present thus suggesting that the general territory of *haber de* is being

overtaken by *tener que*. The increasing use of *tener que*, whether an indirect object is present or not, demonstrates that this form is being used as often in low-focus clauses as in high-focus clauses which seems to point to the grammaticalization of this construction.

### **3.2.8 Tense-Aspect-Mood**

As mentioned in section 2.6.8, tense, mood and aspect play an important role in determining how much an obligation construction has grammaticalized. Recall that, in his study of German modals, Heine (1995:26) indicates that modals are more likely to express deonticity if they are in past or perfect tenses. Similar tendencies are found in studies conducted by Palmer (1990) and Tagliamonte and D'Arcy (2007) on English modality. In their studies on Spanish modality, Sirbu Dumitrescu (1988) and Silva-Corvalán (1995: 90) claim that *tener que*, when used as an obligation marker, tends to occur in present tense. Based on these findings we can hypothesize that the newer forms, especially in the earlier centuries, will be more likely to occur in present tense and, possibly, in the perfect tenses.

Heine (1995) identifies several important criteria that help identify epistemic contexts. One of these criteria implies that the event can not have taken place yet. This would implicate future tense in epistemic contexts. In addition, he emphasizes the importance of probability stating that, although the event is non-factual, there is a high probability that it will occur. Palmer (1990) and Coates (1995; 1983) echo this reference to non-factuality. Of course, non-factuality or probability is related to mood. Given these criteria for epistemicity, we would assume that constructions occurring in conditional or subjunctive forms are more likely to be non-factual and thus related to a more epistemic reading.

Another important tendency in this factor group is habitual action. Both Coates (1983) and Westney (1995) identified habitual uses of English deontic modals in their descriptions of the modal system. Although they did not identify habitual action as a possible point in the grammaticalization path, they did identify specific contexts in which these actions occur. For example, Westney (1995) found that the habitual action must refer to both the present and the future when referred to with *have to*, whereas with *have got to* and *must* the habitual action can only refer to the future. Is it possible that this consistent reference to the future is an indicator of grammaticalization? This might be answered if we see regular patterns among these four constructions.

### **3.2.8.1 *Haber de***

In the 16th century, like in previous results, *haber de* accounts for the majority of the examples, regardless of tense, aspect or mood (Table 3-21). If we look at the tendencies of this marker across centuries it is difficult to see any clear patterns except for the steady decline of the use of *haber de* in every context. One possible pattern may be the less dramatic decline of *haber de* in contexts where it occurs in present tense. However, even in this context, it suffers a dramatic drop in the 20<sup>th</sup> century. There is a slight increase from the 16<sup>th</sup> century to the 17<sup>th</sup> and 18<sup>th</sup> centuries of *haber de* in imperfect (+3%), which is likely to be associated with habitual actions and may tell us something about the path of grammaticalization. We also see an increase of this construction in future tense, which does fulfill one of the criteria outlined in Heine (1995) for epistemicity. There is also an increase in cases of *haber de* in subjunctive up until the 19<sup>th</sup> century, which is associated with non-factuality. Similarly related to non-factuality is the conditional mood, a context in which *haber de* has managed to maintain control through the 20<sup>th</sup> century. The fact that *haber de* still occurs a good

percentage of the time in future tense (35.1%) and conditional (38.1%) suggests that this construction is associated with epistemicity.

Table 3-21. Frequency of *haber de* occurring with the factor group tense-aspect-mood

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
Tense-Aspect-Mood	present	765	88.7	440	86.3	466	77.3	201	32.8
	future	7	70	15	88.2	4	26.7	40	35.1
	preterit	14	87.5	34	75.6	30	38.5	27	11.3
	imperfect	234	91.4	561	94	167	66.5	39	14.4
	present	2	66.7	0	0	0	0	0	0
	perfect								
	conditional	3	75	8	72.7	5	22.7	45	38.1
	present	10	55.6	11	68.8	6	46.2	0	0
	subjunctive								
	imperfect	36	94.7	7	58.3	8	57.1	1	5.9
	subjunctive								
	future	6	100	0	0	0	0	0	0
	subjunctive								
	other	14	70	30	90.9	0	0	1	2.3

### 3.2.8.2 *Tener de*

In absolute numbers this construction is used the most in present tense in the 16<sup>th</sup> century (n=74/89). This could partially be explained by the fact that third-person represents 70% of all contexts during this century. If this is not the case, the tendency to occur more in present tense may limit this construction to deontic readings as Heine (1995) suggests. This trend continues into the 17<sup>th</sup> and 18<sup>th</sup> centuries (n=24/32) before *tener de* falls out of use in the 19<sup>th</sup> century. What is interesting about this is that present tense only accounts for 41% of the data during these two centuries, yet *tener de* still is found most often (75%) in the present tense further supporting claims made by Heine (1995). This is likely another factor for the death of this form.

### 3.2.8.3 *Tener que*

When looking at the frequency of this construction in the 16<sup>th</sup> century we see the majority of the cases are in present tense (n=12), although it accounts for only 1.4% of the present tense data (Table 3-22). In the 17<sup>th</sup> and 18<sup>th</sup> centuries, *tener que* continues

Table 3-22. Frequency of *tener que* occurring with the factor group tense-aspect-mood

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
Tense-Aspect-Mood		N	%	N	%	N	%	N	%
	present	12	1.4	29	5.7	87	14.4	248	40.5
	future	2	20	2	11.8	11	73.3	65	57
	preterit	2	12.5	8	17.8	45	57.7	206	86.2
	imperfect	6	2.3	26	4.4	59	23.5	188	69.6
	present perfect	0	0	1	25	13	100	15	100
	conditional	0	0	2	18.2	16	72.7	63	53.4
	present subjunctive	5	27.8	4	25	5	38.5	13	86.7
	imperfect subjunctive	1	2.6	4	33.3	6	42.9	15	88.2
	future subjunctive	0	0	0	0	0	0	0	0
	pluperfect subjunctive	0	0	0	0	0	0	7	100
	conditional perfect	0	0	0	0	0	0	1	100
	subjunctive pluperfect	0	0	0	0	0	0	1	100
	other	4	20	3	9.1	11	100	0	0

to occur mostly in present tense, even though it is used with a wider variety of other tense, aspect and mood markers during these centuries. This expansion continues into the 19<sup>th</sup> century, particularly its occurrence in preterit (37.7%), present perfect (100%) and future tense (73.3%). By the 20<sup>th</sup> century, the culmination of this trend is that *tener que* practically monopolizes all contexts, with the few exceptions being future tense, present tense and conditional. Again, this clearly indicates that this form is quickly taking over the obligation system and the few remaining areas –future and conditional– will likely take longer given their association with epistemicity and *haber de*. Present tense also seems to be an area where there is still substantial variation, which indicates

that, despite the takeover by *tener que*, the other constructions are still used in contexts with deontic readings.

#### 3.2.8.4 *Haber que*

This construction seems to occur primarily in present tense and imperfect tense, although there are cases of it occurring in other forms as well. This pattern is evident through all centuries even into the 20<sup>th</sup> century where *haber que* maintains a small but stable percentage of the contexts in present tense (26.3%) and imperfect (15.9%). On the one hand, if we consider that the present tense is more associated with deontic readings, we could conclude that *haber que* has not moved into epistemic contexts yet. On the other hand, the fact that this construction is also associated with imperfect tenses could suggest that this construction is associated with habitual actions, which may point to movement along the cline of grammaticalization.

#### 3.2.9 Temporal Adverbs

This factor group was also chosen because of the tendencies found with the habitual use of obligation markers in the English modal system by Coates (1983) and Westney (1995). Aside from using aspect markers like the imperfect tense, temporal adverbial clauses or adverbs can also indicate habitual actions (3. 4). Although there were a minimal amount of temporal adverbs in general, Table 3-23 shows the tendencies throughout the centuries.

- 3.4 *se lo conté a mi confesor, porque **todos los días habíamos de confesarnos*** (González, 1999: CREA)  
'I told it to my confessor, because we would have to confess every day.'

In addition to looking for the tendencies of these markers to appear in contexts that implied habitual activity, all adverbs referring to a future action were also examined.

This is based on the reference to an action not yet completed as being one of the criteria identified by Heine (1995) as characteristic of epistemicity. As a result, it was important to investigate whether future temporal adverbs were part of the grammaticalization process. One way to do this would be to look to see if there is an increase or decrease in the frequency of co-occurrence of future adverbials for each marker. If there appear to be tendencies across the centuries that repeat with each marker, this could be another way to determine how far along the grammaticalization path each construction is.

Table 3-23. The occurrence of all obligation markers with temporal adverbial Markers across centuries

		16 <sup>th</sup>		17 <sup>th</sup> /18 <sup>th</sup>		19 <sup>th</sup>		20 <sup>th</sup>	
		N	%	N	%	N	%	N	%
<i>haber de</i>	habitual	15	83	13	86.6	11	68.7	5	17.8
	future	42	97.6	47	95.9	18	62.1	19	25.6
<i>tener de</i>	habitual	2	11.1	1	6.6	0	0	0	0
	future	0	0	2	4	0	0	0	0
<i>tener que</i>	habitual	1	5.5	1	6.6	4	25	19	67.8
	future	1	2.3	0	0	11	37.9	46	62.1
<i>haber que</i>	habitual	0	0	0	0	1	6.2	4	14.2
	future	0	0	0	0	0	0	9	12.1
Total temporal markers (including non future/habitual)		77		100		68		150	

Looking at Table 3-23, however, we see that there is no clear pattern that the constructions follow. As in the other factor groups, *haber de* accounts for most occurrences with both future and habitual markers during the 15<sup>th</sup> and 16<sup>th</sup> centuries but slowly loses its ground as *tener que* becomes more and more frequent. By the 20<sup>th</sup> century *haber de* only accounts for 17.8% of all contexts with habitual markers compared to *tener que* with 67.8%. A similar result is found with contexts in which future markers occur with *haber de* accounting for 25.6% of the cases in the 20<sup>th</sup> century whereas *tener que* accounts for 62.1%. *Haber que* doesn't seem to occur with any

temporal markers until the 20<sup>th</sup> century with the exception of one case in the 19<sup>th</sup> century.

### 3.3 Multivariate Analyses

Although the results from the frequency data above illustrate that the obligation system in Mexican Spanish has experienced a shift over the last few centuries, they do not give a detailed account regarding which factor groups have the greatest effect on these changes, or where exactly in the grammar these changes are taking place. Multivariate analyses can provide this type of information.

Aside from the numerous studies on changes that happen in modal systems across languages (e.g. Lyons, 1977, Bybee and Pagliuca, 1985, Traugott, 1989), the development of strong obligation in Spanish is largely unknown. Based on the results reported in the previous sections and the dramatic increase in frequency of *tener que* and the subsequent decrease in frequency of *haber de*, we might expect to see similar patterns of change in the conditioning factors across time. That is, the factors that influence the speaker to use one obligation marker over another may also fluctuate over time. The results from the frequency data hint at this, with the large fluctuations in type frequency from one century to another, but results from multivariate analyses can give a more detailed look as to when and where these shifts began.

Given the low frequency of all of the variants except *haber de* during the earlier centuries, it was impossible to conduct multivariate analyses. In effect, it was only possible to look at the variation of these constructions through a multivariate lens during the 19<sup>th</sup> and 20<sup>th</sup> centuries. Even then, it was only possible to compare *haber de* and *tener que* during the 19<sup>th</sup> century. In the 20<sup>th</sup> century, the frequency of these forms was

slightly more balanced, so an analysis could be done to compare each of the remaining three constructions.

Despite the limited number of multivariate analyses, the results distinguish certain patterns in the data that may have otherwise been difficult to ascertain. In fact, the most dramatic changes occurred between the 19<sup>th</sup> and 20<sup>th</sup> centuries, so the snapshot provided for these two time periods reveals the factors that constrain the use of each of these markers in a period of time where a shift appears to be occurring. A detailed look at the results of these analyses is offered.

Each subsection examines and presents the results of the binomial up-and-down multivariate analyses. The first three analyses have *haber de* as the application value, or the measuring source. In section 3.3.4 *tener que* is used as the source of measurement in order to determine the likelihood of a speaker choosing *tener que* or *haber que*. Each subsection contains a table (see Tables 3-27, 3-28, 3-29, 3.30) detailing the results. The column marked 'N of variant' gives the total number of the application variant in that particular context. The next column, labeled 'percentage of variant' presents the relative frequency of occurrence of the application value in the given context. The 'weight' column gives the probability values of the likelihood of the variant occurring in that particular context. A weight of .5 or above indicates that the context favors the use of the application, whereas a value of less than .5 would suggest a less favorable context. The further the number is away from .5, and keeping in mind its relative position among other factors in its group, the stronger the favoring or disfavoring effect is. A weight that is close to .5 indicates that the factor neither favors nor disfavors the variant. Weights for factor groups that are rendered not significant by

the binomial up-and-down are placed in brackets. Finally, the column entitled ‘total number of variant’ provides the total number of both variants that occurred in the given context.

In addition to the multivariate analyses, a comparative analysis of *haber de* and *tener que* across centuries is examined in section 3.4. The comparative analyses give a cross-century comparison of the conditioning factors of *haber de* and *tener de* to see if changes have occurred in terms of the magnitude of effect or the hierarchy of constraints (see section 2.5 for a detailed description on the hierarchy of constraints). The results from this comparison are presented in Table 3-31.

Table 3-31 provides a list of both significant and non-significant factor groups, in order of their magnitude of effect in the 19<sup>th</sup> century. The column labeled ‘direction of effect’ provides a summary of which factor groups were altered in terms of the strength of effect from the 19<sup>th</sup> century to the 20<sup>th</sup> century. The two factor groups that show a change in order, direct objects and verb type, are examined in further detail in order to identify whether changes also occurred within the hierarchy of constraints. Table 3-32, which is provided for this portion of the comparative analyses has two columns labeled ‘weight.’ These are the probability values from each century. The order of these weights is compared. A change in order suggests changes in the underlying grammar of Mexican Spanish regarding these forms, which will be discussed in more detail below.

### **3.3.1 *Haber de* vs. *Tener que*: 19<sup>th</sup> Century**

Although there were enough examples of *tener que* to conduct a multivariate analysis, the type frequency of this form is still quite limited in the 19<sup>th</sup> century. As a result, there were not enough examples to consider type of sentence, preceding clitic or indirect objects in the analysis. The fact that these groups could not be considered

further illustrates the limited contexts within which *tener que* can occur at this time. That being said, a multivariate analysis was carried out with the remaining factor groups, which can be seen in Table 3-24.

There are three significant factor groups: direct objects, verb type and polarity. The direct objects factor group has the strongest magnitude of effect, whereas polarity has the weakest. Within the direct object factor group, it was necessary to combine the objects that were coded as middle voice, intransitive *se* or copular with the category labeled as 'other', due to the limited number of examples and similar preliminary statistical results in each of these groups. This factor slightly favors the use of *haber de*, but this tendency is tenuous at best. Both animate and inanimate objects show no tendency to favor or disfavor *haber de*, whereas oblique complements tend to strongly disfavor the use of *haber de*.

Within the factor group verb type, there was also a need to combine groups given similar preliminary results and the limited number of examples. Copular verbs were combined with psychological/emotion verbs. The verb *hacer* was collapsed with action verbs. Moreover, verbs of influence, perception verbs, and the verbs *haber/tener* were collapsed into the group labeled as 'other.' There are a few notable tendencies within this group. We do see that there is a slight tendency for psychological/emotion and copular verbs to favor the use of *haber de*. This was also found in the results for type frequency. This reaffirms that *tener que* is still not used as often with these types of verbs, which indeed are the types of verbs most often associated with an epistemic reading. There is also a tendency for communication verbs and action verbs to disfavor *haber de*. This goes along with what Cornillie (2007: 219) suggests when he finds that

these types of verbs infer a deontic reading when they occur with *tener que*. Again, this is another indication that *tener que* has not expanded into contexts that include lexical items associated with epistemicity.

Table 3-24. Analysis from the 19th century: Linguistic factors conditioning haber de occurrence in comparison with tener que in CORDE. Probability of Input: 74.7 (.747) , N =686 /939

Factor Group		N of <i>HD</i>	% of <i>HD</i>	Weight	% of data
Direct Objects					
	other	341	81	.56	45
	(in)animate	294	71	.49	44
	oblique	51	51	.29	11
<i>Range</i>				.27	
Verb Type					
	pysch/emotion	239	86	.64	30
	other	118	71	.47	18
	action	281	67	.43	45
	communication	48	66	.40	7
<i>Range</i>				.24	
Polarity					
	negative	135	81	.59	18
	positive	551	71	.48	82
<i>Range</i>				.11	
Grammatical person					
	1 <sup>st</sup> person	267	74	[.52]	38
	other	419	73	[.49]	62
<i>Range</i>					

Log likelihood = -515.642 P value= .05 Total Chi-square = 51.0837 Chi-square/cell = 1.1610\* The brackets [ ] indicate that this factor was not statistically significant.

The last significant factor group, polarity, shows that contexts where negation is present slightly favor *haber de*. This could be an indication that, despite the movement of *haber de* into more epistemic contexts, it has still managed to maintain control in certain contexts, negation being one of them. This may be one of the last areas into which *tener que* expands as it continues to increase in frequency.

### 3.3.2 *Haber de* vs. *Tener que*: 20<sup>th</sup> Century

The multivariate analysis for these two variants in the 20<sup>th</sup> century was more revealing compared to the 19<sup>th</sup> century given the increased amount of occurrences and

the ability to include every factor group in the analysis. As shown in Table 3-25, the following factor groups were significant: verb type, grammatical person, preceding clitic, direct objects, polarity, and type of sentence. Indirect objects were not found to be a significant factor group.

Within the factor group verb type, which had the greatest magnitude of effect, it was necessary to collapse several factors. Verbs of perception and verbs of influence were combined into the group labeled as 'other'. In addition all occurrences of *hacer* were combined with all action verbs. Results indicate that copular verbs strongly favor the use of *haber de* (.76), which corroborates the results discussed in section 3.2.1.2 and suggest that the association of copular verbs with epistemicity has perhaps slowed the expansion of *tener que* into this territory. We see a similar trend with psychological/emotion verbs, although they favor *haber de* to a lesser extent (.62). Action verbs disfavor the use of *haber de* thus supporting the claims that this construction is moving or has moved into primarily epistemic contexts, primarily because Heine (1995) and Cornillie (2007) associate action verbs with deontic readings.

Within the factor group of animacy and grammatical person, contexts with a third-person inanimate subject tend to favor the use of *haber de*. Bybee and Fleischman (1995), Heine (1995) and Pietrandrea (2005) suggest that epistemic contexts are more often found with third-person inanimate subjects. This association could have hindered the expansion of *tener que* into this territory. Contexts with second-person subjects also tend to favor *haber de* slightly. Given that Coates (1983) has identified second-person as the grammatical person most associated with strong obligation, we can conclude that *haber de* still has strong ties to obligation readings. However, based on the fact that

contexts with first-person subjects, another grammatical person strongly associated with obligation, disfavor the use of *haber de*, it is also clear that *haber de* is losing ground.

The presence of a preceding clitic also seems to favor the use of *haber de*. Given that linguists like Myhill (1988) associate clitic climbing with more grammaticalized forms, we would expect that they are more likely found with the older form *haber de*, which is indeed the case.

Several scholars have also identified low-focus clauses as associated with relative clauses, since they often supply background information (Klein-Andreu, 1991; Bybee et al., 1994). These clauses are often identified by negation and inanimate objects. We do not see a pattern either way within the factor group of direct objects with inanimate objects. There is a slight tendency for contexts with animate objects to favor the use of *haber de*, but this is not strong. There is also a tendency for contexts with middle voice and intransitive *se* to favor the use of *haber de* (.66). The nature of middle voice makes the object behave as both the subject and object, which may be why these contexts are associated with the older construction. Thus it could be suggested that middle voice is also associated with epistemicity although more research is needed to verify this.

We see that, within the factor group polarity, contexts with negation slightly favor the use of *haber de*. Although the aforementioned linguists associate negation with low-focus, it is also identified by Coates (1983), Westney (1995) and Silva-Corvalán (1995) as occurring less often in epistemic contexts, given the limited contexts within which negation can occur with epistemic forms. Taking this into consideration, it is possible that *haber de* has grammaticalized and is thus associated with low-focus clauses, which gives it an association with negation. However, as it moves into more

epistemic clauses this association should decline, which has not happened, thus indicating that this form has not fully grammaticalized.

Table 3-25. Data from the 20th Century: Linguistic factors conditioning *haber de* occurrence in comparison with *tener que* in CORDE. Probability of Input: 26.7 (.267), N = 354/1219

Factor Group		N of <i>haber de</i>	% of <i>haber de</i>	Weight	% of data
<i>Verb Type</i>					
	copular	71	56	.76	10
	psych/emotion	80	39	.62	17
	other	68	29	.51	19
	communication	29	24	.46	10
	action	106	20	.39	44
				37	
<i>Range</i>					
<i>Grammatical Person</i>					
	3 <sup>rd</sup> inanimate	85	49	.66	14
	2 <sup>nd</sup> person	51	39	.57	11
	3 <sup>rd</sup> animate	158	26	.49	49
	1 <sup>st</sup> person	60	19	.39	26
				26	
<i>Range</i>					
<i>Preceding Clitic</i>					
	yes	67	46	.70	12
	no	287	27	.47	88
				23	
<i>Range</i>					
<i>Direct Objects</i>					
	middle voice	29	46	.62	6
	other	86	34	.59	22
	animate	42	32	.57	12
	inanimate	92	22	.46	37
	oblique	50	20	.41	23
				21	
<i>Range</i>					
<i>Polarity</i>					
	negative	45	41	.66	9
	positive	309	28	.48	91
				18	
<i>Range</i>					
<i>Type of Sentence</i>					
	interrogative	36	43	.66	7
	declarative	318	28	.49	93
				17	
<i>Range</i>					
<i>Indirect Objects</i>					
	present	84	30	[.54]	23
	not present	270	29	[.49]	77

Log likelihood = -650.39 P value = .05 Total Chi-square = 345.437 Chi-square/cell = 1.1326 \*The brackets [ ] indicate that this factor was not statistically significant.

The final significant factor group is the type of sentence. There is a slight tendency for interrogative sentences to favor *haber de* (.66). Coates (1983), Bybee and

Fleischman (1995), and Tagliamonte and D'Arcy (2007) suggest that epistemic contexts limit the possibility of interrogative sentences. The assumption was, then, that if one of the markers had moved into primarily epistemic contexts it would less likely be found in interrogative sentences. Since this is not the case with *haber de*, we assume that this construction still has a stronghold in contexts with deontic readings.

### 3.3.3 *Haber de* vs. *Haber que*: 20<sup>th</sup> Century

The vast majority of cases with *haber que* occur in third-person and, if a clitic is present, it unfailingly occurs after the infinitive verb. As a result of these limitations, grammatical person and preceding clitic were excluded from the multivariate analyses. In addition, due to the limited number of occurrences of *haber que* in interrogative sentences, type of sentence was also excluded.

There are two significant factor groups, direct objects and verb type (Table 3-26). The results of direct objects, which had the greatest effect on the choice of variants, show that contexts with inanimate objects disfavor the use of *haber de*. Given that a similar trend was found when comparing *haber de* with *tener que*, we can assume that these types of contexts are areas in the grammar into which the newer constructions are expanding. There is also a slight tendency for the objects labeled as 'other', which includes copular objects, middle voice and intransitive *se*, to favor the use of *haber de*. This is likely due to the stronghold *haber de* seems to have on contexts with copular verbs, and verbs used with middle voice. This is somewhat corroborated in the statistics of the second significant factor group, verb type.

Within verb type, there were several factors that had to be combined. Verbs of communication, influence, perception and the verbs *haber* and *tener* were combined with the group labeled as 'other.' Contexts with the verb *hacer* were also collapsed with

action verbs, while contexts with copular verbs were combined with psychological/emotion verbs. In the results, we see that copular and psychological/emotion verbs slightly favor the use of *haber de*, which is similar to what was found in the comparison between *haber de* and *tener que*. This further supports the idea that *haber de* has a stronghold in this area of the grammar. There is also a tendency for action verbs to disfavor the use of *haber de*, signifying the possible development of *haber de* into non-deontic associated contexts.

Table 3-26. Data from the 20<sup>th</sup> century: Linguistic factors conditioning *haber de* occurrence in comparison with *haber que* in CORDE. Probability of Input: 61.3 (.613) , N =354 /586

Factor Group		N of <i>HD</i>	% of <i>HD</i>	Weight	% of data
Direct Objects					
	other	170	73	.62	40
	animate	42	64	.55	11
	oblique	50	56	.45	15
	inanimate	92	46	.37	34
<i>Range</i>				25	
Verb Type					
	pysch/emotion	138	71	.58	33
	other	110	65	.55	29
	action	106	48	.39	38
<i>Range</i>				18	
Polarity					
	positive	309	61	[.51]	86
	negative	45	56	[.45]	14
Indirect Objects					
	not present	270	58	[.50]	75
	present	84	58	[.49]	25

Log likelihood = -368.426 P value= .05 Total Chi-square = 44.6078 Chi-square/cell = .9913 \*The brackets [ ] indicate that this factor was not statistically significant.

### 3.3.4 *Tener que* vs. *Haber que*: 20<sup>th</sup> Century

Similar to the previous analysis, grammatical person and preceding clitic are not considered in the multivariate analysis. In addition, due to the limited number of examples of *haber que* in interrogative sentences, this group was also excluded. The significant factor groups were polarity and direct objects (Table 3-27).

Table 3-27. Data from the 20<sup>th</sup> century: Linguistic factors conditioning *tener que* occurrence in comparison with *haber que* in CORDE. Probability of Input: 79.4 (.794), N =865 /1097

Factor Group		N of HD	% of HD	Weight	% of data
Polarity	positive	799	80	.52	91
	negative	66	65	.32	9
				20	
Range Direct Object	oblique	200	84	.59	22
	other	254	80	.52	29
	inanimate	411	76	.45	49
				14	
Range Verb Type	other	263	82	[.55]	29
	action	427	79	[.50]	50
	psych/emotion	175	75	[.44]	21
Indirect Objects	not present	665	80	[.51]	76
	present	200	77	[.46]	24

Log likelihood = -553.905 P value= .05 Total Chi-square = 41.67 Chi-square/cell = 1.157 \*The brackets [ ] indicate that this factor was not statistically significant.

Within the factor group polarity there is a strong tendency for contexts where negation is expressed to disfavor the use of *tener que*. In the previous analyses, we also saw that these same contexts disfavored the use of *haber de*. This suggests that this is an area in the grammar into which *haber que* was able to generalize.

The other significant factor group shows a slight tendency for contexts with oblique complements to favor the use of *tener que*. The factor labeled as ‘other,’ which includes middle voice, intransitive *se* and copular objects doesn’t show a tendency one way or the other. This is probably because neither one of these two constructions is strongly associated with these types of objects yet, since they are still more likely to be used with *haber de*. Contexts with inanimate objects slightly disfavor the use of *tener que*, although the tendency is too slight to draw any concrete conclusions.

### 3.4 Comparative Analysis

Based on the fact that the only common multivariate analyses possible in both centuries were the comparisons of *haber de* and *tener que*, these were the only analyses that could subsequently be compared using the sociolinguistic comparative method (see section 2.6 for details on this method). As a reminder, a shift in the order of effect of significant factor groups across centuries suggests a possible change in progress. If in addition to the shift in significant factor group order, there is also a change in the constraint hierarchy within the factor groups, this also indicates a change and suggests where in the grammar the change is probably occurring. In the case of the present study a change in the constraint hierarchy will indicate where in the grammar the changes are taking place. The results of this comparison are seen in Table 3-28. Factor groups that were significant in the 20<sup>th</sup> century but were excluded in the 19<sup>th</sup> century were not included in this comparative analysis.

Table 3-28. Variation of *haber de* and *tener que* across two centuries

Factor Group	19 <sup>th</sup> Century	Direction of effect	20 <sup>th</sup> century
Direct Objects	significant	different	significant
Verb Types	significant	different	significant
Polarity	significant	same	significant
Grammatical Person	not significant	--	significant

As we can see, there is a shift in the order of significant factor groups from the 19<sup>th</sup> to the 20<sup>th</sup> centuries. In the 19<sup>th</sup> century, direct objects had the greatest magnitude of effect whereas in the 20<sup>th</sup> century, verb type has the greatest effect. This suggests possible changes in progress and warrants further examination of the constraint hierarchies of each of these two factor groups.

An examination of the constraint hierarchies of each factor group exposes several important changes. Within the factor group verb type, there is a change in the order of

action verbs and communication verbs. This could be an area in the grammar where change is occurring. Although the details are more ambiguous in the direct object factor group, it appears as if there has been no change in the order of constraints. This would suggest that the change that is occurring is centered on verb type. As these forms grammaticalize and expand it is likely that there will be a shift in the conditioning effects. This Table (3-29) demonstrates that these shifts, particularly when looking at *haber de* and *tener que*, are likely to happen within these two factor groups.

Table 3-29. Analysis of constrain hierarchies of *haber de* vs. *tener que* across the 19th and 20th centuries.

19 <sup>th</sup> Century Factor Group		Weight	20 <sup>th</sup> Century Factor Group		Weight
Direct Objects			Direct Objects		
	middle/copular/ other	.56		middle	.62
	inanimate	.49		other	.59
	animate			animate	.57
	oblique	.29		inanimate	.46
<i>Range</i>		27	<i>Range</i>		21
Verb Type			Verb Type		
	Copular	.64		copular	.76
	psych/emotion			pysch/emotion	.62
	other	.47		other	.51
	action	.43		action	.39
	communication	.40		communication	.46
<i>Range</i>		24			37

In sum, the results from the written corpus suggest that specific factors are associated with the older, more frequent forms. This suggests that these factors (e.g. copular verbs, existential subjects, preceding clitics) are likely connected to movement towards epistemic contexts. Accordingly in the earlier centuries, the newer constructions occur often with certain factors (e.g. communication verbs, oblique objects), which indicates that these factors may help determine the age of each of these constructions. These results will be discussed in more detail in Chapter 4.

### 3.5 DLNE and Oral Corpus

In general, the results of the oral corpus suggest an ongoing change is occurring in Mexican Spanish deontic modality, as can be seen in Table 3-33. Speakers overwhelmingly select *tener que* as the preferred obligation marker in natural speech (92%). The findings also show an overall tendency for these obligation markers to not appear in future tense (.3%) suggesting future action may already be encoded in the meaning of the construction, which in effect could indicate that this marker may have advanced along the grammaticalization cline towards more epistemic contexts. However, considering that the future tense (both periphrastic and synthetic futures were considered) only occurred one time in all of the data (3.55), it is difficult to conclude that this is the result of advancement of these markers along the grammaticalization cline.

This could be explained by the nature of obligation to imply future meaning through intention, which would mean that a future marker would be unnecessary because future is already implied (Traugott, 1989; Bybee et al., 1994; Aaron, 2007). It could also be a reflection of the type of questions that the participants were asked. Most questions were narrative in nature and often involved talking about stories about past experiences. Whatever the reason, it is difficult to make any conclusive statements about the absence of the future tense in the corpus.

- 3.5 *Si estar con mi familia porque digo en un momento si pasara el examen voy a tener que salir* (Roberto, 8/2010)  
'Yes, be with my family because if I do pass my exams I am going to have to leave.'

Within verb type the only variation that is seen occurs with verbs that fall into the other category where there are 6 examples of *haber que* (10.3%) and 2 of *haber de* (3.4%) as can be seen in Table 3-30. *Haber de* is not common but, of the 8 times it

occurs, 3 are with psychological/emotional verbs, 1 with copular verbs and 1 with the verb *tener*. This means 5 of the 8 occurrences (63%) appear in contexts associated with epistemicity. However, when these contexts are compared to the total number of occurrences of psychosocial/emotional verbs, copular verbs or with the verb *tener*, *haber de* only accounts for a minimal percentage (8%) of these cases. *Haber que* does not occur with psychological/emotional verbs nor with copular, which again suggests that this form may not be grammaticalizing.

Table 3-30. Data from the 21<sup>st</sup> Century: Frequency of occurrence of each obligation marker within factor group verb type

Verb Type	HD		TQ		HY		Total	
	N	%	N	%	N	%	N	%
action	1	.5	176	94.6	9	4.8	186	51.5
<i>hacer</i>	0	0	30	100	0	0	30	8.3
perception	0	0	2	66.7	1	33.3	3	.8
communication	1	4.3	18	78.3	4	17.4	23	6.4
psychological	3	33.3	6	66.7	0	0	9	2.5
<i>haber/tener</i>	1	3.8	24	92.3	1	3.8	26	7.2
copular	1	4.3	22	95.7	0	0	23	6.4
influence	0	0	6	100	0	0	6	1.7
other	1	1.8	48	87.3	6	10.9	55	15.2

What is equally worth mentioning is that *haber que* is relatively absent in both the DLNE and the oral corpus. It only accounts for 5.8% of the data in the oral corpus. A similar percentage is seen in most of the centuries in the DLNE as well: 19<sup>th</sup> century (5.3%), 18<sup>th</sup> century (0%), 17<sup>th</sup> century (6.8%) and in the 16<sup>th</sup> century (2.5%). Based on these data *haber que* seems stagnant in terms of frequency. This corroborates previous results, which indicate that this form does not appear to be moving towards epistemicity. Nevertheless, it is important to note that in both the spoken and the DLNE corpora, the context typically includes personal narratives, which could essentially create a bias towards the other constructions since *haber que* is used with non-referential subjects the majority of the time.

Within the group of grammatical person and animacy we see that 100% of the cases with existential subjects occur with *tener que* (Table 3-31). This is a departure from the results found in the written corpus. This is an indication that obligation is grammaticalizing at a quicker pace in oral speech (which is expected as writing is more conservative) and thus *tener que* has already taken over the system and is moving into epistemic contexts and slowly taking over that territory as well. A cursory glance at the DLNE corpus provides little support, considering that the only examples that contain existential subjects occur in the 16<sup>th</sup> and 17<sup>th</sup> centuries.

Also worth noting are the three cases of *haber que* referring to first-person (3.56) identified as non-third-person through the use of the intransitive *se* or some reflexive object pronoun. We can see in example 3-56 that the object *nos* is expressing the middle voice. That is, the object is the same as the subject. In order for this to be the case, however, the construction *hay que* cannot be impersonal since *nos* refers to the first-person plural. Taking that into consideration, this example clearly illustrates that *hay que* is not exclusively used as an impersonal expression. Nevertheless, examples of this construction in first-person are relatively rare compared to the total number of examples in first-person (3.1%) but it does seem to parallel a similar pattern found in the written corpus.

- 3.6 *Había un enfermo que no había medicamento y **hay que** cooperarnos entre todos y darle a esa persona.* (Marisol, 7/2010)  
'If there was a sick person and no medicine, we had to cooperate among ourselves and provide for this person.'

Given that it occurs only in the 19<sup>th</sup> and 20<sup>th</sup> centuries in the written and the oral data and accounts for 3 of the 21 cases (14%) of *haber que* in the oral corpus, this could point to the beginning of this form moving into new contexts with different

grammatical persons thus opening the door towards expansion and grammaticalization. In terms of sociolinguistic variables there was very little difference between the choices of variants, which was expected. For example younger speakers used *haber de* and *haber que* with the same frequency as the older speakers. Similar results were found when comparing male to female speakers. What this suggests is that, if a change has occurred and *tener que* has virtually taken over the system, this change already had taken place and had been adopted into this community by the time the younger generation began speaking. Regardless of the reason, this community seems to rely heavily on *tener que* as an obligation marker.

Table 3-31. Data from the 21<sup>st</sup> Century: Frequency of occurrence of each obligation marker according to factor group

		HD		TQ		HY		Total	
		N	%	N	%	N	%	N	%
Grammatical Person	1 <sup>st</sup> person	1	.8	123	96.9	3	2.4	127	35.2
	2 <sup>nd</sup> person	1	2.1	47	97.9	0	0	48	13.3
	3 <sup>rd</sup> person animate	5	3	142	86.1	18	10.9	165	45.7
	3 <sup>rd</sup> person inanimate	1	12.5	7	87.5	0	0	8	2.2
	existential subject	0	0	13	100	0	0	13	3.6

The fact that *tener que* accounts for the majority of the cases also echoes reports of similar results found in the English obligation system. In their study on Toronto English, Tagliamonte and D'Arcy (2007) find that *have to* has taken over the entire obligation system. Given that such results are being found in other languages, this offers support to the idea that results found in this oral corpus may in fact reflect the status of the obligation system in modern Mexican Spanish.

Taking into consideration that 62% of the examples in the oral corpus are in present tense (Table 3-32), we would assume that all of these constructions would be employed since Silva-Corvalán (1995), Heine (1995) and others suggest that present

tense is associated with a deontic reading. This is not the case, however, since *tener que* is the marker of choice. This further supports the hypothesis that *tener que* has taken over the obligation system. This expansion indicates that this form has grammaticalized and spread into new contexts in a relatively short amount of time, possibly within 100 years.

Another tendency found in the data is that 23% of the total examples are in the imperfect tense. As previously mentioned, one of the ways habitual actions are identified is through aspect. *Tener que* accounts for all but 2 of the cases of imperfect verb form. If we compare this to the 18<sup>th</sup> and 19<sup>th</sup> centuries from the DLNE corpus, where imperfect accounts for 50% of the cases and *tener que* is only found in 23% (N=11) of those cases, we can see that there appears to be a shift. This could be another identifiable way to determine if an obligation marker is grammaticalizing. This will be discussed in more detail in Chapter 4.

Table 3-32. Data from the 21<sup>st</sup> Century: Frequency of occurrence of each obligation marker within factor group tense-aspect-mood (TAM)

		HD		TQ		HY		Total	
		N	%	N	%	N	%	N	%
TAM	present	6	2.4	228	89.4	21	8.2	255	70.6
	preterit	0	0	14	100	0	0	14	3.9
	future	0	0	1	100	0	0	1	.3
	imperfect	2	2.4	82	97.6	0	0	84	23.3
	conditional	0	0	1	100	0	0	1	.3
	subjunctive	0	0	3	100	0	0	3	.8
	other	0	0	3	100	0	0	3	.8

Due to the small number of examples in both of these corpora, it was difficult to draw any concrete conclusions about the tendencies found. It is nevertheless important to examine these data, given that the tendencies found in both corpora seem to follow similar patterns found in the written corpora. This is further support that these constructions were grammaticalizing and, in fact, continue along this path. As a result of

the minimal data available, only the data extracted from the oral corpus was discussed in detail, although reference to the DLNE corpus was made throughout the section to corroborate (or not) possible tendencies.

### 3.6 Summary

This chapter has covered the changing tendencies of each of these obligation markers. Through a careful examination of text and type frequency, we verified that the obligation system in Spanish has indeed undergone a shift since the 16<sup>th</sup> century. The results suggest certain patterns of change, that may help corroborate the claims of universal paths of grammaticalization in modality. The discussion of these tendencies is examined in more detail in Chapter 4.

In addition to analyzing patterns of frequency, multivariate analyses were conducted for the 19<sup>th</sup> and 20<sup>th</sup> centuries. The results further implicate a change in the system, particularly within the factor group verb type. These data provided a snapshot of the patterns of use among three obligation markers in the 20<sup>th</sup> century. The multivariate analyses for each variant further suggest that *tener que* is taking over obligation. We see a similar trend in the oral and DLNE data.

The oral data are almost exclusively comprised of examples of *tener que*. Despite the patterns of use identified in the written data, this corpus offers a different picture of the obligation system, showing minimal use of *haber que*, decreasing use of *haber de* and a takeover by *tener que*. Although there may be alternate explanations for these results, which will be discussed in Chapter 4 and Chapter 5, the results from the DLNE corpus seem to suggest a similar pattern. Also, the minimal number of examples in both the DLNE and oral corpus, and the subsequent restrictions on the conclusions that can be drawn from these two corpora, are also discussed in detail in Chapter 4.

## CHAPTER 4 DATA ANALYSIS

Based on the overall frequency patterns of these four obligation markers, the system appears to change in a way similar to that of obligation systems in other languages. That is, diachronically, these forms appear to undergo what Hopper and Traugott (1993) and Tagliamonte and D'Arcy (2007) term as longitudinal renewal, which is essentially a new way to say the same thing. Bybee (2003) explains that this often happens when the older forms become so frequent that they lose their intensity and force. As a result, new, innovative forms enter the system and begin to compete with the older forms (Hopper and Traugott, 2003: 125).

In Figure 4.1 we see that, in the early centuries, *haber de* is the primary obligation marker used. This continues through the 17<sup>th</sup> and 18<sup>th</sup> centuries, although by this point *tener que* has entered the system and is slowly starting to expand. The expansion of *tener que* correlates with the loss of territory of *haber de*, both in overall frequency and within the various semantic contexts. For example, as *tener que* is used more and more often with stative verbs *haber de* starts to be used less and less in the same context. This suggests that these two constructions may be ultimately filling similar semantic roles, albeit at different periods in time. The question that remains, though, is how *tener de* and *haber que* play a part. At a quick glance, we can see that the increase in frequency of use of *tener que* and *haber que* also correspond to the decrease in frequency of *tener de*. So the question is not whether *tener de* and *haber que* were part of this longitudinal renewal or not, but rather why they weren't sufficient enough to replace *haber de*. This was answered by looking at the development of each of these

markers to determine the ‘motivations and enabling factors’ behind the rise and fall of each construction (Hopper & Traugott, 2003: 71).

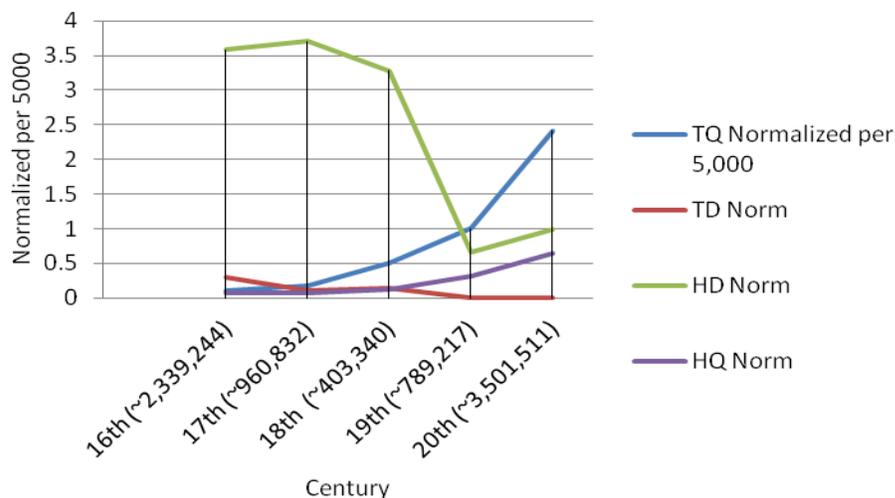


Figure 4-1. Frequency of variants across centuries

In addition to understanding the story behind each of these constructions, the results reported in Chapter 3 also shed light on the path of grammaticalization of obligation markers in Spanish. Essentially, each of these forms reflects a different stage of grammaticalization. *Haber de* is the oldest form dating back to the 12<sup>th</sup> century and well established as the marker of choice by the 13<sup>th</sup> century (Yllera, 1980:99). *Haber que* came into the system a little later and, around the 14<sup>th</sup> century, became limited paradigmatically, occurring only with 3<sup>rd</sup> person impersonal subjects (Cornillie, 2007:227). *Tener de*, according to Yllera (1980:110-116), was used as a deontic construction as early as the 13<sup>th</sup> century. *Tener que* also appeared around the 13<sup>th</sup> century as an obligation marker, although it becomes more frequent around the 16<sup>th</sup> century (Yllera, 1980: 117). Taking the age of each of these markers into consideration and the difference between them in terms of when they were used most frequently, the

stages of each of these constructions throughout the centuries reveal a relatively clear path of change.

The following sections will review the results presented in Chapter 3 and situate these obligation markers in their respective paths of grammaticalization. The discussion will begin with section 4.2, which will review the various phases in grammaticalization pertinent to the present study. Section 4.3 will expand on this path by re-examining how subjectification plays a role in the development of these constructions. Then, in the sections that follow, the cumulative results of each marker will be discussed and the various factors that place these markers along points in the grammaticalization path will be identified. Section 4.4 will include a general discussion about what the tendencies of each of these constructions represent when compared against the other variants.

#### **4.1 Identifying the Path of Change**

One of the primary goals of this study was to answer several central questions in regards to the aforementioned markers of obligation. These questions were presented in Chapter 1 but are listed here again for review.

- What are the semantic differences among these four obligation markers?
- What are the factors conditioning the use of each of these obligation markers.
- Have these forms continued to grammaticalize?
- If so, what is the path of grammaticalization of each marker?
- Have they also started to undergo subjectification?
- Does the path of change of these markers corroborate claims made for universal paths of grammaticalization in modality?

I have shown in this study that these questions can be answered by examining the path that each of these markers takes between the 16<sup>th</sup> and 21<sup>st</sup> centuries. In order to do this, however, it was paramount to operationalize the relevant stages in

grammaticalization and subjectification. By doing this, the extent to which these forms have developed and advanced along this path could be measured.

#### 4.1.1 Grammaticalization

Recall that each of these constructions has already been identified as having gone through some of the processes identified in grammaticalization. Grammaticalization, the gradual process whereby a construction becomes less specific and moves towards more grammatical, abstract meanings is identified by several characteristics. To review, we know that these constructions evolved from the transitive verbs expressing possession, *haber* and *tener*, into semi-auxiliaries expressing obligation. They have lost many of the properties that were associated with the transitive possession verb, which in grammaticalization is identified as decategorialization.

We are also confident in assuming that these verbs have become bound to the preposition *de* or the conjunction *que*, which is another stage in grammaticalization. In the data, in fact, there were not any examples of these constructions occurring with a clitic or any other element between the verb and *de/que* after the 16<sup>th</sup> century thus indicating a high level of fusion between these two elements. These constructions became even more fused as time progressed. After the 16<sup>th</sup> century there weren't any examples of a word, such a clitic, occurring between the constructions and the infinitive. It was determined that this is further support that these forms have already entered the grammaticalization process, even as obligation markers.

As these verbs became more and more fused to *de/que* we expected to see autonomization take place, which often includes the divergence of these newer constructions from the source form. We demonstrated that there was early evidence for this, as was discussed in Chapter 1. Both *tener que/de* and *haber de/que* are no longer

associated with the notion of possession. In fact, the possession meaning has been all but lost with the verb *haber* in the written and spoken corpora. Again, this was a clear indication that these forms had already started to grammaticalize.

Another important indicator of grammaticalization is the presence of layering or when two or more forms appear to compete for the same territory, another possible sign that grammaticalization is taking place. We know that these forms were in direct competition by the 16<sup>th</sup> century because they emerged in what appeared to be similar contexts, such as in the examples seen here in (4.1-4.4)

- 4.1 *no **tengo que decir** más de remitirme a lo que ellos dijeron* (Cortés, Hernán, 1524: CORDE)  
'I don't have to say anything more than referring to what they said'
- 4.2 *De las demás casas de este signo no **hay que dezir** más de lo que está dicho atrás.* (Sahagún, Fray Bernardino de, 1576-1577: CORDE)  
'It's unnecessary to say anything more than has already been said about houses of this type'
- 4.3 *á quien **tengo de dezir** verdad y lo que siento* (Moya y Contreras, Pedro, 1575: CORDE)  
'to whom I have to tell the tell the truth and what I feel'
- 4.4 *que ninguna falta haze en lo que **ha de decir*** (Sahagún, Fray Bernardino de, 1576-1577: CORDE)  
'nothing is lacking in what I have to say'

In order to further trace this path, some of the processes that supported the notion of movement along this path were operationalized. For example, retention of source semantic content was one of the processes in grammaticalization operationalized through verb type (Bybee et al., 1994, Torres Cacoullós, 2011). This was done to see what kind of verbs with which each of these constructions is associated. It was expected that the infinitive verbs that follow these obligation constructions may in fact be similar to

those verbs associated with these constructions prior to their movement into the obligation system.

In addition to retaining lexical content, these constructions also generalize as they grammaticalize. This generalization, or semantic bleaching, is easily measured and analyzed through relative frequencies. As these constructions are used more frequently, they begin to occur in new contexts. Semantic bleaching can be seen in every factor group but it is particularly noticeable with the verb, both in type and lexicon. The data show that generalization in meaning correlates with movement towards more subjective readings.

As strong obligation markers expand into new territories, they can often infer an intention or future meaning (Cornillie, 2007; Bybee & Fleischman, 1995; Bybee et al., 1994, Fleischman, 1982) prior to moving into epistemic contexts. Taking this into consideration, it was determined that intention needed to be operationalized as another possible way to determine the age of these markers. Since this has been operationalized via temporal adverbs we can expect that the beginning stages of this inference will likely occur with adverbs or adverbial clauses referencing the future (Blas Arroyo's, 2008; Torres and Walker, 2008; Aaron, 2006). As Hopper and Traugott (2003:82) mention, however, the intention meaning must occur often in order for it to have an effect. As a result, we expected that an increase in occurrence with temporal adverbs will likely indicate movement towards epistemicity.

Inference, another important process in grammaticalization, was also operationalized through several different factor groups: sentence type, polarity, verb type and grammatical person. In regards to sentence type, as was discussed earlier, if

the sentence was a declarative sentence, it was less likely to infer epistemicity since there are limitations when it comes to epistemicity and interrogative sentences (Tagliamonte & D'Arcy, 2007; Cornillie, 2007; Silva Corvalán, 1995; Coates, 1983). Thus, the forms that occur less often in interrogative sentences are more likely to infer epistemicity.

Polarity, another factor dealing with inference, has been found to occur in epistemic contexts in limited cases (Olbertz, 1998; Heine, 1995; Silva-Corvalán, 1995; Palmer, 1990; Coates, 1983). This would imply that the constructions that do not occur with negation are more likely to have epistemic meanings. In fact, Olbertz (1998:405) goes so far as to say that 'strong truth commitment can be expressed in positive terms only' and suggests that sentences such as 4.5 are not possible.

4.5    *No **tiene que ser** fácil para ti.* (Olbertz, 1998: 405)  
      'It does not have to be easy for you.'

Grammatical person and animacy, also a factor group associated with inference, is an important element in determining whether a context can infer epistemicity. As Pietrandrea (2005), Bybee and Fleischman (1985) and Coates (1983) mention, epistemicity is often associated with third person subjects. Cornillie (2007), Heine (1995) and Coates (1983) also suggest that animacy plays a role in determining whether a context will have an epistemic reading as well. Finally, Wärensby (2004) identifies existential subjects as being uniquely epistemic. We posited, then, that the constructions associated with third-person, inanimate subjects are more likely to infer epistemicity whereas constructions that are employed more often with first- and second-person subjects will be less likely to infer epistemicity (Krug, 2000).

Mood is yet another factor that may infer epistemicity. Given that linguists such as Coates (1983), Palmer (1990) and Heine (1995) categorize an epistemic proposition as non-factual (where the probability of the proposition occurring is high), we can connect non-factuality to mood based on its non-factual status. Therefore, constructions commonly occurring in conditional or subjunctive mood are more apt to infer epistemicity given the scale of non-factuality inherent to these types of morphological markers.

The last factor group associated with inference is verb type. We know that obligation markers are more likely to express deonticity when they occur with communication verbs or action verbs (Cornillie, 2007). Conversely, the constructions are more likely to infer epistemicity if they occur with stative verbs. The assumption was, then, that as these obligation markers become more and more grammaticalized they would occur more often with stative verbs, and less often with action verbs and communication verbs.

Another process identified in research on grammaticalization is the movement of the grammaticalizing form from high focus clauses into low focus clauses. Based on previous research, low focus clauses are typically associated with more grammaticalized forms (e.g. Diaz-Campos, 2011; Bybee et al., 1994; Klein Andreu, 1991). Based on this research, low- and high-focus clauses were deemed an important area in the grammar to help determine how much the obligation markers have grammaticalized. This process was operationalized through object pronouns (via animacy), polarity and verb type. We know that low focus clauses are associated with negation, inanimate subjects or objects and stative verbs. As a result, we postulated

that, as these obligation markers increase in frequency, the occurrence with the aforementioned factors will be indicators of movement along the grammaticalization path. On the other hand, if some of these constructions are not used in contexts with the previously mentioned factors, this could be an indicator that the construction is not grammaticalizing.

The grammaticalization of obligation markers also includes the shift of these constructions from an agent-oriented reading to a more speaker-oriented reading. This sub-process of grammaticalization was also considered when examining the results of each marker.

#### **4.1.2 Subjectification**

An integral part of the grammaticalization path of obligation markers is movement towards more speaker-oriented contexts, also known as subjectification. Accordingly, determining the extent of subjectification that each of these forms has undergone, particularly in terms of weak or strong epistemicity, is paramount. Recall that Traugott (1989) identifies three tendencies (see Chapter 1) she thinks are part of the subjectification process. She proposes a path of subjectification specific to agent-oriented modals (e.g. obligation). That path is shown in Table 1-1 with each tendency (1-3) noted above where she believes it occurs. Langacker (1999a, b) defines subjectification in a similar to Traugott (1989) but focuses on synchrony and the attenuation of the modal in terms of the amount of control the agent has over the proposition. Both the stages identified by Traugott (1989), as well as the synchronic attenuation found in Langacker (1999a,b), have been operationalized in a number of different studies, as mentioned in Chapter 1. Unfortunately, the large majority of these studies were not relevant to the variants under examination in this study.

Given that so little research has attempted to operationalize subjectification, several factors were chosen that were thought to reflect this process. The factor groups associated with subjectification are i) tense and aspect, ii) temporal adverbs, iii) verb type and iv) grammatical person/animacy. These factors primarily centered on the stages identified by Traugott (1989) although the characteristics of weak epistemicity correlate with Langacker's (1999a,b) idea of synchronic attenuation as well. Traugott (1989) posits that many deontic constructions have already passed through two stages of subjectification when they undergo the shift from being main verbs to deontic verbs. Considering that each of these markers has already made that shift, we concluded that they have already passed through the first two stages of subjectification and started the path towards the third stage of subjectification.

Traugott (1989: 39-40) labels the first phase of the third tendency as weak epistemicity, which correlates with what Bybee et al. (1994) refer to as intention. Traugott (1989: 40) identifies weak epistemicity through markers that indicate a typical or habitual action or implicate prophetic or relative future. She describes relative future as referencing a 'later time' but nothing specific like, for example, *later* and claims that the relative future shows that a construction has not fully subjectified.

I deduced that any specificity given to the time period means that the form has advanced along the subjectification path. This was operationalized through temporal adverbs. If there was reference to the future with a specific time adverbial marker we assumed that this form had entered into strong epistemicity, whereas if the future marker was non-specific we assumed weak epistemicity. We can also look at tense as a way to measure movement into weak or strong epistemicity since Bybee (1987:5 as

cited in Traugott, 1989) suggests that present tense may be a strong indication of movement into future tense.

Habitual actions were measured through the incorporation of aspect and temporal adverbs. We presumed that the use of a temporal adverb connoting a habitual action would likely place the construction in the weak epistemicity phase. In regards to aspect, we know that Traugott (1989) postulates that a deontic obligation marker referring to a typical or habitual action is associated with weak epistemicity, which would imply that constructions commonly found in imperfect or progressive are less likely to infer strong epistemicity.

The other factors that are associated with subjectification –verb type and grammatical person- follow the same tendencies that were discussed in 4.1. That is, as these constructions undergo subjectification they are more likely to occur with stative verbs. In addition, we also postulated that contexts with third-person inanimate subjects, as well as existential subjects, are more likely to have epistemic and thus subjective meanings than those in first- and second-person.

## **4.2 Data Analysis**

In order to determine the path of change for the obligation system in Spanish, a thorough examination of the tendencies is paramount. Only then can we draw more general conclusions and identify the possible evidence that implicates where on the grammaticalization path each of these constructions is. In the following sub-sections, we analyze the various tendencies presented in Chapter 3 in each factor group. These tendencies are connected with the aforementioned evidence that has been shown to account for stages in the grammaticalization process. Additionally, tendencies that are found to be unique to a particular marker, or not identified in previous research, are also

discussed. These patterns may show that Spanish has a distinct path of grammaticalization for obligation markers.

#### 4.2.1 Tener de

When there are many choices in a language for a specific function, like obligation, sometimes the choice of forms is reduced, such as in the case of *tener de*. This often happens because the other constructions start to generalize and move into new contexts. Specialization, as it's sometimes referred to, is particularly relevant when we talk about the obligation system in Spanish and of the disappearance of *tener de*, although sometimes, as Hopper and Traugott (2003:115-116) point out, 'old forms may continue to coexist therefore specialization does not necessarily entail the elimination of alternatives, but may be manifested simply as textual preferences, conditioned by semantic types, sociolinguistic contexts, discourse genres, and other factors.' We see that in the case of *haber que*, *haber de* and *tener que* they continue to coexist. In the case of *tener de*, however, it was eliminated. As Figure 4.1 above indicates, *tener de* essentially stops being used as a marker of obligation around the end of the 18<sup>th</sup> century. Although there are an isolated number of cases of this marker occurring in the later centuries, these are few and far between. The big question then, is why did this happen?

One of the first signs of trouble for this construction is found within the factor group verb type. Although in general, *tener de* was used occasionally with all verb types, in the 16<sup>th</sup> century it was primarily associated with the verb *hacer* and with communication verbs. Remember that *tener de* was frequently found with the verbs *hablar* 'speak' and *librar* 'to free' during the 14<sup>th</sup> century (Yllera, 1980: 113) so communication verbs and action verbs may have been the entry point for this construction. Between the 16<sup>th</sup> and

the 18<sup>th</sup> centuries *haber de* expands and essentially takes over contexts with communication verbs, accounting for 76% of these contexts, essentially wiping out any hold *tener de* may have had in this area. In addition, because *tener que* is presumed to be associated with *tener de* (according to Cornillie, 2007), communication verbs and action verbs were also the likely entry point for *tener que* and, ultimately, more competition for *tener de*. *Haber que* also started to appear with communication verbs early on so between the three constructions *tener de* quickly lost the association with communication verbs.

During the 16<sup>th</sup> century *tener de* was often used in contexts with psychological and emotion verbs, an area in the grammar where epistemicity could possibly be inferred. This was a potential opportunity for *tener de* to expand into more subjective contexts. However, like with communication verbs, both *tener que* and *haber que* used these same contexts as vehicles for their expansions. As a result, *tener de* loses all these initial associations by the end of the 18<sup>th</sup> century. In sum, the factor group verb type appears to be an area in the grammar where *tener de* lost a lot of ground quickly.

We see that, in terms of grammatical person, *tener de* was strongly associated with deontic contexts in the 16<sup>th</sup> century, accounting for 50% of all contexts with first-person subjects. In addition, it was also frequently employed with third-person animate subjects, where it accounted for 39% of all *tener de* occurrences. Both these results suggest that *tener de* was well established as a deontic modal. But by the 17<sup>th</sup> and 18<sup>th</sup> centuries we see that *haber de* increased by more than 35% in contexts with first-person subjects. In fact, both *tener que* and *tener de* lost territory to *haber de* during this period, although *tener que* eventually gained it back. Even though *tener de*

did not gain this territory back, it did manage to retain some of its association with these contexts as the second most frequent construction used with first-person subjects during the 17<sup>th</sup> and 18<sup>th</sup> centuries. This indicates that this area in the grammar was likely one of the last areas to lose its association with *tener de* before this construction fell out of use.

The other important trend regarding *tener de* during these centuries is the expansion of *tener que* into contexts with third-person subjects. This development essentially separated *tener de* from the only other grammatical person with which it was associated. And to exacerbate the departure of *tener de*, *haber que* was limited to contexts with third-person animate subjects during these early centuries and also experienced a rise in frequency, thus helping take over the little territory left for *tener de*. We observed a similar pattern in the factor group sentence type where *haber de* increased in frequency with both interrogative and declarative sentences while the new forms *haber que* and *tener que* also experienced increases although they were minimal. *Tener de* can't compete with all three forms, and thus loses ground.

Despite the fact that *tener de* did not expand during the 17<sup>th</sup> and 18<sup>th</sup> centuries and that the majority of the contexts in which it was found were primarily associated with deonticity (e.g first-person subjects, communication and action (*hacer*) verbs), there were still signs that it may expand into territories associated with epistemicity. Early on, it did occur often with psychological and emotion verbs and occasionally with copular verbs. We also saw that *tener de* was used most with verbs in present tense between the 16<sup>th</sup> (83%) and the 18<sup>th</sup> centuries (75%) and typically associated more with positive

polarity than negative polarity. These are all contexts that may have inferred epistemicity, as we see in 4.6.

- 4.6 *que la tierra adonde se da el cacao tiene de ser muy buena; y porque este cacao es comida* (Toribio de Benavente, 1536 – 1541: CORDE)  
'the ground where the cacao is planted must be good; because it's food.'

This evidence led us to believe that this construction would continue to grammaticalize. The one thing this marker did not do as time progressed, however, is increase in frequency and spread into new contexts. In fact, it was quite the opposite. In the 16<sup>th</sup> century it occurred with specific verbs regularly like *comer* (n=9), *ser* (n=5) or the verb *hacer* (n=11) among others. By the 17<sup>th</sup> and 18<sup>th</sup> century it rarely occurred with the same verb more than once, with the exception of the verbs *acabar* (n=3), *dar* (n=5) and *pasar* (n= 3). In addition, it dropped dramatically in frequency. In the 16<sup>th</sup> century it accounted for 7% of all data, but by the 17<sup>th</sup> and 18<sup>th</sup> centuries it only accounted for 3% of all data. So despite the patterns this marker shows towards being associated with weak epistemicity (e.g. contexts in present tense, imperfect and 11% of habitual temporal markers), its decrease in frequency and failure to spread into new contexts did not allow this to happen.

One possible reason to explain why *tener de* appears in so many contexts early on could be based on the similarities this construction shared with *haber de*. Yllera (1980: 111) suggests that *tener de* came into being as a result of the substitution of *tener* for *aver* 'have' in contexts historically reserved for *aver*. It is possible that this substitution may have spilled over into the periphrastic uses of *haber de* as well. If this were the case, then it would explain why *tener de* was employed in so many different contexts from the beginning.

Yllera (1980:97) also mentions that, in old Spanish, constructions with verb+*de*+infinitive often indicated purpose or finality. She credits this formulaic construction for the quick rise in frequency of *haber de*. It could also be another possible reason why *tener de* was found in so many contexts, despite being a new construction. If we consider that *haber de* was already associated with finality and purpose this may have easily associated the construction with intention thus propelling this form into contexts implying future actions. As a way to differentiate between intention and obligation perhaps another construction, not containing *de*, was needed. The need for an obligation marker removed from the finality or purpose feature of the *de* construction could also explain the appearance of *tener que* and the rise in frequency of *haber que*. This idea supports the theory proposed by Cornillie (2007) who suggests that *tener que* came from *tener de* and not from the relative clause *tener+object+que+verb*.

Another important aspect to consider when we look at the disappearance of this construction is to look beyond the variable context. Given that this marker still retained possession readings in certain contexts, as illustrated in example 4.7, it was important to determine whether the decrease in frequency of *tener de* as an obligation marker was paralleled by a corresponding increase in frequency by the possession construction *tener de*.

4.7 *Deseos **tengo de** verlos enfadados* (Fernández de Lizardi, 1818: CORDE)  
'I have the desire to see you (pl) annoyed.'

As we can see in Figure 4.2, there were 16 examples of this construction in the data being used to express possession (15% of total data) in the 16<sup>th</sup> century. In the 17<sup>th</sup> and 18<sup>th</sup> centuries there were 52 examples (60% of total data) and in the 19<sup>th</sup> century there were 10 instances (83%). The rise in frequency of cases where this construction

expressed possession correlates to the decline in frequency of the construction being used as an obligation marker. Taking this correlation into consideration, it is likely that the contexts where possession was expressed also played a principle role in the decline and eventual disappearance of *tener de*.

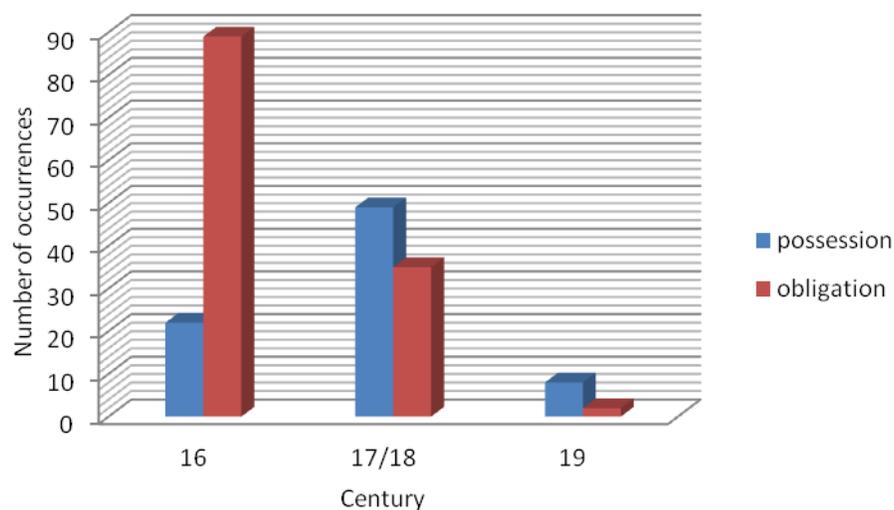


Figure 4-2. Comparison of *tener de* as an possession marker and as an obligation marker across centuries

#### 4.2.2 Tener que

One of the first patterns we observe with the construction *tener que* is that it started out being used with a variety of verbs, but was especially used with communication verbs, which might suggest that it may be closely related to *tener de*. Recall that this is based on research conducted by Yllera (1980: 113) who finds that *tener de* is most often found with the verb *hablar* in the earlier centuries. Again, this observation offers further support that *tener que* may be derived from *tener de* since it was found in similar contexts at this point in time.

If it is the case that *tener que* entered the language on the tail of *tener de*, it may explain why we also see that this construction starts to appear in contexts that could

possibly infer epistemicity. In the 16<sup>th</sup> century, for example, it had already started to be used with psychological and emotion verbs. This trend continued into the 17<sup>th</sup> and 18<sup>th</sup> centuries where it accounted for 14% of all contexts with psychological and emotion verbs. Likewise, within the factor group sentence type, *tener que* is not associated with interrogative sentences until the 20<sup>th</sup> century, meaning that it often occurred in contexts that linguists (e.g. Coates, 1983; Corvalán, 1995) typically associated with epistemicity in the research. It was also noted that, as early as the 17<sup>th</sup> and 18<sup>th</sup> centuries, *tener que* occurred more often with inanimate subjects than with animate subjects, again suggesting that this form was occurring in contexts with possible epistemic readings from early on.

Recall that Ruiz de Mendoza Ibáñez and Pérez Hernández (2004: 106) suggest that *tener que* is more associated with internal obligation. They also suggest that this association disappears in the past tense but claim it is preserved in the future tense (Ruiz de Mendoza Ibáñez and Pérez Hernández 2003: 110). Given that internal obligation is associated with epistemicity, we deduced that when *tener que* occurs in past tense it is less likely to be epistemic, whereas when it occurs in the future or present tense it is more likely to be epistemic. This is supported by Solano-Araya (1982: 17) who suggests that ‘epistemic modals have limitations when it comes to [...] past tense.’. If we look at the data, we see that in the 16<sup>th</sup> century *tener que* occurred in the preterit in about 13% of the cases. As time progressed, this number steadily increased until the 20<sup>th</sup> century, at which point *tener que* accounted for 86% of all cases in preterit. We see a similar increase in frequency of *tener que* appearing in future tense as well. It

appears that, as time progressed, *tener que* became increasingly used in both deontic and epistemic contexts.

We know that Traugott (1989) also finds that tense plays a role in subjectification although she separates relative future from deictic future. When we look at the factor group tense-aspect-mood we see that *tener que* accounted for 70% of all contexts in imperfect in the 20<sup>th</sup> century. Considering that the imperfect often indicates habitual actions, this was identified as corresponding with weak epistemicity (Traugott, 1989). This tendency was corroborated by the percentage of habitual temporal markers that occurred with *tener que* during the same century (68%). Both the frequency of occurrence of *tener que* in the imperfect, along with the high number of temporal markers indicating habitual actions, are clear indicators that this form was associated with weak epistemic during the 20<sup>th</sup> century.

We also see that this construction accounted for 57% of the contexts in future tense during the 20<sup>th</sup> century. Traugott (1989) identifies this as another possible characteristic of subjectification. *Tener que* accounted for 68% of the temporal future markers during the 20<sup>th</sup> century, of which about half (n=24) were non-specific and the other half (n=22) were specific. Considering that Traugott (1989) regards deictic future as an indication of strong epistemicity, we can conclude that *tener que* had already moved into contexts associated with strong epistemic readings in the 20<sup>th</sup> century.

When *tener que* began to expand into all of the territories in the 19<sup>th</sup> century *haber de* managed to maintain its association with preceding clitics, accounting for 99% of all contexts with a preceding clitic during the 19<sup>th</sup> century. This seems to be one of the areas that was a stronghold for *haber de*. This is corroborated by the multivariate

analyses where, in both centuries, preceding clitic still strongly favors the use of *haber de*. However, by the 20<sup>th</sup> century *tener que* did start to appear in these contexts at a rapid rate. In the span of just one century, *tener que* accounted for 54% of the cases with preceding clitics. Based on the claims made by Myhill (1988) and Gutiérrez (2010) who suggest that clitic climbing occurs more often with auxiliaries than with other verbs, coupled with the observation that *tener que* comprised 54% of the cases with a preceding clitic during the 20<sup>th</sup> century, we can deduce that this marker progressed further along the grammaticalization path during this century.

Many of these tendencies are substantiated by the multivariate analyses. For example, in both the 19<sup>th</sup> and 20<sup>th</sup> centuries, when compared to *haber de*, we see that *tener que* continues to be linked to communication and action verbs by the disfavoring effect these two factors have with *haber de*. We also observe that, although *tener que* was used with psychological and emotion verbs early on, this factor still favors the use of *haber de*, supporting the claim that *tener que* is still in the earlier stages of subjectification. We do see that, in the 20<sup>th</sup> century, interrogative sentences slightly favor the use of *haber de*. This could be one of the areas that allowed *tener que* to be associated with epistemicity in this later century. We also see in the 20<sup>th</sup> century that there is a slight tendency for inanimate objects to disfavor the use of *tener que*. Again, like with interrogative sentences, this is an opportunity for *tener que* to occur in contexts that infer epistemicity.

Overall, there is an observable renewal taking place when we look at the frequency patterns of *tener que* in comparison to *haber de*. *Tener que* appears to be used in the same contexts as *haber de*, just at a later time period. In fact, according to

the spoken data from the 21<sup>st</sup> century *tener que* has literally taken over the obligation system. This raises the question as to whether *haber de* will continue on the grammaticalization path or, like *tener de*, fall out of use. As Fernández de Castro (1999) mentions, *haber de* is already primarily limited to written or formal registers, which points to the possible disappearance of this construction. The descent of *haber de* is discussed in more detail in the next section.

#### 4.2.3 Haber de

One of the first tendencies we notice about *haber de* is that its overall frequency decreases over time. As has been mentioned on several occasions, *haber de* had control over the entire obligation system during the 16<sup>th</sup> century. This is seen in every factor group. However, as time progresses we see a general decrease in frequency across the board.

Within the verb type factor group *haber de* continued to regularly occur with the psychological and emotion verbs, and the verbs *haber* and *tener* throughout the centuries. Considering that these verbs are all perfect areas in which epistemic readings can be inferred, we assumed that this construction was moving towards epistemicity. This was corroborated by the fact that *haber de* was also most likely to occur with third-person inanimate subjects and existential subjects. Interrogative sentences, however, did not support this assumption. During the 20<sup>th</sup> century *haber de* was often used in interrogative sentences (40%) whereas it was only used with declarative sentences 23% of the time. We also observe that when negation was expressed *haber de* was used 31% and was used only 24% of the time when negation was not expressed. Considering that negation is associated with low focus clauses, but it is not

associated with epistemicity, we might say that this result indicated that *haber de* was in an intermediate stage of grammaticalization and moving towards epistemicity.

If we look at the results of the factor group polarity, we see that *haber de* was associated with low-focus clauses. Klein Andreu (1991) mentions that stative verbs are also indicative of low-focus clauses, an area in which *haber de* has had control since the 16<sup>th</sup> century. This was also supported by the fact that *haber de* represented more of the total percentage of contexts where negation was present (31%) than those where it was not (24%). Again, although negation corresponds to low focus clauses, it is not associated strongly with epistemicity.

We see that within the factor group indirect objects *haber de* occurred as much with an indirect object as it did without (24%). In turn, where direct objects were represented *haber de* occurred with animate objects 27% of the time and inanimate objects only 18% of the time. Although this suggested that this form was still primarily relegated to high focus clauses, we know that epistemicity is associated with stative verbs and, logically, stative objects, an area where *haber de* accounted for 46% of the data. Overall, we see that *haber de* did have characteristics (negation, stative verbs, no preference with IO) of occurring in low focus clauses.

Although it has been well documented in the literature that *haber de* expresses future (e.g. Cornillie, 2007; Yllera, 1980), we see that, based on its occurrence with temporal adverbs referencing the future, this tendency has been declining. In the 16<sup>th</sup> century *haber de* accounted for 98% of all contexts with future adverbials. During the 20<sup>th</sup> century, however, it only accounted for 26%. Although, of the 19 future adverbial markers with which it occurred, 47% (n=9) of those adverbs make reference to specific

time periods. This indicates that this construction was still associated with and/or moving towards strong epistemicity during the 20<sup>th</sup> century. This was corroborated by tendencies found in the tense-aspect-mood factor group.

*Haber de* occurred in 38% of all cases when conditional mood was employed and 35% when future was used. This only strengthened the association with epistemicity. We also observe that it accounted for 33% of the contexts in present tense, which has also been found to be associated with future reference (Bybee, 1987; Traugott, 1989). In addition to relative future, habitual actions are associated with weak epistemicity as well. We see that in the 20<sup>th</sup> century *haber de* accounted for 14% of those cases, which was more than a 50% decrease from the previous century, thus indicating movement from weak to strong epistemicity.

The multivariate analyses echo what the frequency data tell us. For example, in the 19<sup>th</sup> century, when compared to *tener que* we observe that direct objects were a significant factor group and had the highest magnitude of effect, followed closely by verb type. Both of the factor groups pointed to *haber de* moving into epistemic contexts. For example, the factor labeled 'other' within direct objects slightly favors the use of *haber de*. This is expected since this factor included copular objects and these types of objects are typically associated with epistemicity. We see the same pattern with verb type where copular verbs and psychological/emotion verbs favored the use of *haber de*<sup>1</sup>.

---

<sup>1</sup> And we also have the same confusing result with negation, where it favored the use of HD which associated it with low focus clauses but disassociated it with epistemic contexts according to (authors).

In the 20<sup>th</sup> century the same analysis showed even more robust results and supporting evidence that *haber de* was the more likely construction to be used in epistemic contexts. Verb type was the factor group with the greatest effect and, like in the 19<sup>th</sup> century, copulas and psychological/emotion verbs still favored *haber de*. We also see that grammatical person was significant in this century and, not surprisingly, we see that the third-person inanimate subjects favored the use of *haber de*, again associating this construction with epistemicity. Contexts with preceding clitics strongly favored *haber de*, which pointed to *haber de* being a more grammaticalized construction than the other constructions. In addition, contexts with copular objects favored *haber de* as did contexts with negation present. An interesting result that we did not see in the 19<sup>th</sup> century analysis was that interrogative sentences favored the use of *haber de*. This may be an indication that, even though this form was moving into epistemic contexts, it still retained its deontic reading as well. When we compare *haber de* with *haber que* the results, once again, indicate that copular objects and copular/psychological/stative verbs favor the use of *haber de*.

#### **4.2.4 Haber que**

One of the first indicators of language change, as suggested above, is semantic bleaching. We see semantic bleaching occurring with *haber que*, particularly in the 19<sup>th</sup> and 20<sup>th</sup> centuries, when it started to occur with a wider variety of verb types. Taking a more detailed look at the data, we see that this construction occurred mostly with the verb *decir* in the 16<sup>th</sup> century (n=9/21), which illustrates the limitations of this marker at that point in time. This also illustrates that this form retained some of its lexical roots given that Yllera (1980:109) finds that *haber que*, prior to being used an obligation marker, often occurred with the verbs *decir* and *hablar* (among others). But, even as

early as the 17<sup>th</sup> century, *haber que* had started to spread into new contexts. This was illustrated by the number of new verbs with which this construction was appearing during this century. This is similar in grammatical person where, in the 19<sup>th</sup> century, we start to see contexts where there are first- and second-person subjects.

One of the primary tendencies we see with this construction in regards to grammatical person is its expansion into new territories. In fact, this expansion is still rare enough in the data that calling it a tendency may be too strong. Nevertheless, there are examples in all of the data sets of *haber que* being used in a context that implies a subject other than an impersonal one. Being confined to contexts with impersonal subjects, along with the inability for *haber que* to occur with preceding clitics, led us to believe that this construction may be doomed to deonticity. This was further supported by studies such as Cornillie et al. (2009) who highlight the inability of *haber que* to occur in any context but deontic in their data. That being said, we did find examples of this marker in the corpora in contexts with first- or second-person subjects (ex. 3.54, repeated here in 4.8) suggesting that this limitation could possibly change. Although *haber que* did not occur in contexts inferring epistemicity in the data, (ultimately supporting claims made by Cornillie et al., 2009) a cursory glance at the corpus CREA produced a context where epistemicity can be inferred (4.9) with *haber que*.

- 4.8 *Los podremos ayudar para que puedan hacer alguna carrera. Total, no **hay que** precipitarnos...* (Azuela, 1985: CREA)  
'We will be able to help them so that they can run some race. Anyway, we don't have to rush.'
- 4.9 *pero la verdad es que **hay que** ser un hombre rarísimo para echarle la culpa de todo a tu abuelito.* (Bryce Echenique, Alfredo; 1986: CREA)  
'but the truth is that you have to be a strange person to blame everything on your grandfather.'

Taking this into consideration, along with the results demonstrating that *haber que* has been used in contexts with subjects other than 3<sup>rd</sup> person impersonal, this may be an area of expansion for the construction.

*Haber que* has been used rarely in interrogative sentences throughout the centuries. Knowing that epistemicity is rarely expressed in interrogative sentences, this tendency points to *haber que* occurring in contexts that could be associated with epistemicity. The tendency for *haber que* to not be employed in interrogative sentences, in conjunction with its proclivity to spread into new contexts as time progresses, hints at the possible grammaticalization of this form and movement towards epistemic contexts. This was not corroborated, however, by the results found with the factor group negation. *Haber que* accounted for a higher percentage of occurrences in contexts where negation occurs than those where it does not in every century. Despite the fact that *haber que* occurred in more contexts with positive polarity when considering absolute frequency, the overall percentages illustrated that this construction was used more often with negation.

In regards to tense-aspect-mood, the results indicated that this construction occurred primarily in present and imperfect tense during most centuries. There were also cases of it occurring in conditional and present subjunctive early on, and these did increase during the 20<sup>th</sup> century. We also observe that, in the 19<sup>th</sup> century, *haber que* started to occur in the past tense, which is not associated with epistemicity. However, since imperfect is associated with weak epistemicity (Traugott, 1989) and present tense is often linked to future tense (Traugott, 1989) we concluded that *haber que* might infer weak epistemicity in these contexts. This is supported by the expansion of *haber que*

into new territories during the last couple centuries, particularly into those associated with epistemicity. This is further corroborated by examining the occurrence of this construction with temporal adverbs.

Between the 16<sup>th</sup> and the 19<sup>th</sup> centuries *haber que* did not appear with habitual or future temporal markers. In the 20<sup>th</sup> century however, it appeared with both. The habitual markers suggest connections to weak epistemicity while the future markers suggest an association with strong epistemicity, provided they refer to a specific point in time. Upon further examination of these future markers we see that, of the nine markers that refer to future, seven (78%) refer to a non-specific time in the future, as can be seen in 4.10, which occurs with the non-specific marker *luego* 'later'. This implies that this marker is more likely associated with weak epistemicity than strong epistemicity.

- 4.10 *y luego **había que** tumbar la puerta porque no iba a caber* (Elizondo Elizondo, Ricardo, 1987, CREA)  
'and later one had to knock down the door because it was not going to fit.'

Another interesting result that was seen in the analysis of *haber que* is the relative frequency patterns in the different corpora. In the written corpus the rate of *haber que* increased almost 10 times the rate of occurrence between the 16<sup>th</sup> and the 20<sup>th</sup> centuries. It only represented approximately 2% of all obligation markers during the 16<sup>th</sup> through the 17<sup>th</sup> centuries but had jumped to 16% by the 20<sup>th</sup> century. This was different in the DLNE corpus where *haber que* accounted for about 5% of the data through every century and only 6% of the data in the spoken corpus. This difference seems to highlight a possible preference of *haber que* in the written corpora. If this is indeed the case, the possibility of this marker grammaticalizing will likely diminish.

In the multivariate analysis with *haber de* we see that animate objects favored *haber de* as opposed to *haber que* during the 20<sup>th</sup> century. This again supports a possible semantic shift for *haber que* towards more low focus clauses. This is further supported in the multivariate analysis with *tener que* where inanimate objects and negative polarity disfavored the use of *tener que*. Both of these, indicators of low focus, suggest again that *haber que* may be moving into contexts where weak epistemicity can be inferred. Nevertheless, we also know that Coates (1983), Westney (1983) and Silva-Corvalán (1995) mention that there are limitations to the type of contexts in which epistemicity can be expressed when negation is present. We can assume then that, although *haber que* is moving in the direction of strong epistemicity, it is still in the nascent stages.

In the oral corpus we see that *haber que* is only employed occasionally (6%) and typically occurs with communication verbs and perception verbs. Although we can not directly compare the oral data to the written data, we can still assume that the oral data is likely years ahead of the written data. If this is indeed that case, *haber que* may never move into epistemic contexts as it appears to be stagnant in the spoken data. That being said, some of the same patterns that were seen in the written corpus appear in the spoken corpus as well. For example, we see examples of *haber que* with subjects other than third-person impersonal and all the examples of *haber que* are in present tense.

### **4.3 General Discussion**

Taking all the aforementioned conclusions into consideration there are some observable overall trends that we see in the system that may indicate general trends of obligation markers in Spanish. Additionally, these trends offer insight and support to

postulations put forth about universal paths of grammaticalization and the movement of obligation markers towards more speaker-oriented subjective contexts.

One of the notable trends we see with *tener de*, *tener que* and *haber que* is that all these constructions enter the obligation system being associated with communication verbs. Historically, two of these constructions tended to occur with the communication verb *hablar* before they came to mean obligation (Yllera, 1980). Although we do not know if this is also the way *haber de* entered the system we can conclude that, in modern Spanish, obligation markers enter the system via communication verbs. This may be the beginning of the path.

We also see a similar pattern with action verbs in the earlier centuries, although by the 17<sup>th</sup> and 18<sup>th</sup> centuries these markers begin to appear in other contexts on a regular basis. This may be an indication that, as these constructions move towards epistemicity, they associate less with communication and action verbs. We see some change happening in this area in the comparative analysis between *tener que* and *haber de*. There is a change in the constraint hierarchy between communication and action verbs with verb type. Although *tener que* has clearly not shifted away from its lexical origins and the tendency to appear with these types of verbs, the comparative analyses suggest an impending change. Regardless, we can still postulate that the path from possession to obligation is partially facilitated through action and communication verbs.

We also see a similar pattern with grammatical person where the entry point seems to occur in contexts with both first and second person animate subjects, although given the limitations in which *haber que* can occur, it's hard to draw a concrete conclusion about whether or not these two areas are truly access points. We could also

observe that these constructions were typically found in imperfect or present tense. Although this is not represented in the overall percentage during the 16<sup>th</sup> century, we can see this trend in the absolute frequency of *tener que*, *tener de*, and *haber que*. What is also evident in the data is that these constructions likely entered the language in declarative sentences without indirect objects. The rest of the factor groups do not show general trends during this period thus we can conclude that, while each construction was used in distinct ways, they also entered into the obligation system via a similar path.

One of the early signs of the grammaticalization of these constructions, aside from generalizing and spreading into new contexts, is the tendency to occur with stative verbs, particularly those expressing a psychological state or emotion. This is evidenced by *tener de*, *tener que* and *haber que* early on and, as time progresses, *haber que* and *tener de* cease to appear in these contexts, offering further support for the ongoing grammaticalization of *tener que*, and the apparent halt in grammaticalization of *haber que*. It may be that psychological and emotion verbs are one of the first contexts in which these markers have the opportunity to infer epistemicity.

Another observable general trend found in the data is the general dearth of examples of the markers with negation as these constructions develop. *Haber de* dominates these contexts until the 19<sup>th</sup> century. This suggests that, despite its limitations in epistemic contexts, negation may in fact be an indication of grammaticalization. If this is indeed the case, it would explain the tendency of *haber que* to not appear in contexts with negation and back up the claim that the development of

this construction has slowed since the 16<sup>th</sup> century. What this may point to is another way to determine the age of these forms.

It may be that these markers only become associated with negation when they begin to move into low focus clauses. As they continue to develop and start to become associated with epistemicity, they once again move away from negation. We see this with *haber de* in the 20<sup>th</sup> century where it drops dramatically in frequency in contexts with negation. In addition, in the multivariate analyses polarity is a significant factor group in the 19<sup>th</sup> and 20<sup>th</sup> centuries when comparing *haber de* and *tener que* thus suggesting that this is an area that conditions the choice of marker. The results in the multivariate analysis between *haber de* and *haber que* in the 20<sup>th</sup> century suggest a similar conclusion.

Based on the assumption that *haber que* did not grammaticalize much between the 16<sup>th</sup> and 19<sup>th</sup> centuries, we wouldn't expect it to occur in low-focus clauses, where it would be associated with negation. Conversely, *haber de* seems to have grammaticalized a great deal since the 16<sup>th</sup> century given the aforementioned indications of its movement into epistemic contexts. Thus, we would not expect *haber de* to occur often with negation given the limitations of negation and epistemicity mentioned by Coates (1983), Westney (1995) and Silva-Corvalán (1995). In sum, both of these markers should show little preference when it comes to polarity since they both tend to associate with positive polarity. The multivariate analysis supports these assumptions illustrating that when negation is present there is no preference between these two markers.

The results indicate that these constructions are used in both negative and positive polarity contexts more or less equally and therefore there is no preference. We could postulate that this is an indication that *haber que* is slowly moving into low focus clauses so we see that, occasionally, it is used with negation. Similarly, we can claim that *haber de* is slowly moving into epistemic clauses and thus is associated less and less with negation.

The claim that negation is an intermediate stage of grammaticalization is further supported when we look at the multivariate analysis between *tener que* and *haber que*. We see that negation strongly favors the use of *haber que*. Assuming that *haber que* is slowly moving into low focus clauses, this result substantiates that movement. But additionally, and based on earlier discussed results, we know that *tener que* is showing indications of moving into strong epistemicity (e.g. stative verbs, existential subjects). We would assume then that, as *tener que* moves into strong epistemicity, it will be used less and less with negation, which is reflected in this multivariate analysis.

In sum, we can postulate that these constructions start out not being associated with negation. They develop, move into low focus clauses and begin to occur with negation. As they continue to grammaticalize and move into epistemic contexts they stop occurring with negation. Thus negation is an intermediary stage of the grammaticalization of these obligation markers.

Another general trend we see in the data is the tendency for these newer constructions to retain and maintain some of their pre-obligation lexical features. Recall that Yllera (1980) found that *tener de* and *haber que* were commonly associated with the verb *hablar*. We saw that each of these constructions, along with *tener que*, entered

the obligation system via these types of verbs. We also observe that, in the 20<sup>th</sup> century, there were still tendencies for *haber que* and *tener que* to occur with these types of verbs. This may be an area in the grammar that these constructions will maintain until a new obligation marker enters the system.

As these markers grammaticalize and move into low focus clauses and epistemic contexts, once again we see evidence that low focus clauses may be associated with the intermediate stages of grammaticalization (like with negation). In both the multivariate analyses and the comparative analysis, we see that direct objects are considered a significant factor group. Recall that low focus clauses are often identified by the occurrence with negation and inanimate objects. When we look at direct objects in more detail we see some obvious tendencies. Of course, one of the first tendencies we observe is that in the earlier centuries the newer constructions are used most often when a direct object isn't present. Any suggestion of movement into low-focus clauses occurs in the 17<sup>th</sup> and 18<sup>th</sup> centuries with the construction *tener que* as it starts to increasingly be used with inanimate objects. *Haber que* follows suit in the 20<sup>th</sup> century where we see a 15% increase in contexts with inanimate subjects. In the 20<sup>th</sup> century, as *tener que* expanded into new territory, we found contexts with inanimate objects disfavored *haber de*, suggesting that, at this point, *tener que* was associated with these contexts. This indicates that the occurrence with inanimate objects is an indication of movement into low-focus clauses but that low-focus clauses are more of an intermediary stage that occurs as these markers move into epistemicity.

We see a similar pattern with oblique complements where, as *tener que* generalizes and spreads into new contexts, it begins to occur more often with obliques.

In fact, that multivariate analysis from the 19<sup>th</sup> century indicated that oblique complements strongly disfavored the use of *haber de* when compared with *tener que*. A similar result is found in the analysis between *haber de* and *haber que* in the 20<sup>th</sup> century, where both obliques and inanimate objects disfavor the use of *haber de*, supporting the claim that, as these markers develop, these two contexts are one of the first they acquire. Although the reason for occurring with inanimate objects can be explained by the movement from high-focus clauses into low-focus clauses, the reason why oblique objects are also an area associated with this path is unclear.

Finally, based on the tendency for *haber de* (and later *tener que*) to occur with copular objects, we can also conclude that contexts with these types of objects are one of the last areas these markers acquire as they move towards epistemicity. This may be why this factor group has a change in effect in the comparative analyses. *Tener que* may be edging towards more contexts with copular objects and away from contexts with oblique objects. We can thus postulate a path in terms of occurrence with objects, which appears to be oblique>inanimate>copular.

Other areas in the grammar that seem to occur more towards the end of the obligation path are found in various factor groups. For example, we see that *haber de* continues to occur in contexts with copular verbs into the 20<sup>th</sup> century and, although the newer markers appear with other statives (e.g. psychological/emotion verbs), copulas appear to be among the last contexts in which they occur before they move into epistemic contexts. This would support claims made by Pietrandrea (2005); Cornillie, 2007; and Coates (1983).

The placement of a clitic also seems to be an area that could indicate the age of these constructions. *Haber de* occurs in 99% of the contexts with a preceding clitic as late as the 19<sup>th</sup> century. It isn't until the 20<sup>th</sup> century that *tener que* starts to appear in these contexts and, even so, the contexts with a preceding clitic still strongly favor the use of *haber de*. This is a clear indication that the occurrence with a preceding clitic is an advanced stage of the grammaticalization process of these obligation markers. This corroborates research conducted on clitic climbing that suggests that clitic climbing occurs more often with auxiliary or grammaticalized constructions (Gutiérrez, 2008; Myhill, 1988).

Finally, another important general trend we see in the data is found in the multivariate analysis between *haber de* and *tener que*. The results suggest that this system is not only undergoing change, but that it is in the process of longitudinal renewal. The fact that, in the 20<sup>th</sup> century, the multivariate analysis between *haber de* and *tener que* produces results in which every factor group except indirect objects is significant points to a shift in the system across the board. So despite the path that we identified above, it is also important to recognize that *tener que* has gone through the grammaticalization process at a faster pace than *haber que* and has essentially caught up to where *haber de* was in the 16<sup>th</sup> century by the 20<sup>th</sup> century. As *tener que* continues to move into the territory of *haber de* we would expect a similar trend to occur, which is exactly what we see in the spoken corpus.

As we mentioned in section 3.5, the results from the spoken corpus indicate that *tener que* has essentially taken over the obligation system. Taking into consideration that there have not been studies that look at the diachronic development of obligation

markers in Spanish, it is difficult to determine the lag time, if any, between the written and spoken corpora. Research that has compared written and spoken corpora has had varying results (e.g. Biber, 1995). Based on the relative frequencies of these constructions, however, it appears that the lag seems to vary from century to century.

If we compare the relative frequencies of the DLNE corpus to the written corpus, the 16<sup>th</sup> century written data appear to be about 200 years behind the DLNE corpus. This gap diminishes, however, in the 19<sup>th</sup> century where the normalized frequencies are similar across both corpora, with the exception of *haber de*. The 20<sup>th</sup> century we see a steady increase of *tener que* in the written corpora but the DLNE corpus ends during the 19<sup>th</sup> century so there is no data with which to compare the written corpus. The oral corpus, as we mentioned, shows a sharp increase in the use of *tener que* and a corresponding sharp decrease in *haber de*, while *haber que* seems to have leveled off (when compared to the 19<sup>th</sup> century DLNE). Taking these tendencies into consideration, it is difficult to determine the lag time but it is likely to be no more than 100 years based on the comparisons of earlier centuries.

Regardless of the size of the gap in time, the spoken data clearly show that *tener que* has taken over the system. This, again, supports the idea that obligation in Spanish has undergone longitudinal renewal and *tener que* has essentially replaced *haber de* in most contexts.

## CHAPTER 5 FUTURE DIRECTIONS AND CONCLUSIONS

### 5.1 Future Directions

This paper has provided a detailed account of the development of obligation in Mexican Spanish. One of the principle results is that *tener que* has taken over the obligation system over the last century. This shift from *haber de* to *tener que* is primarily restricted to the 19<sup>th</sup> and 20<sup>th</sup> century, although semantic generalization does begin to occur as early as the 16<sup>th</sup> century. Nevertheless, an important direction in which to go in the future would be to include more spoken data. Comparing these markers across dialects and generations is likely to give a clearer picture of the obligation system today. Recall that Labov (1972) indicates that synchronic variation is often a reflection of diachronic change so detecting changes in a modern system is essentially a window into the past.

In addition to including more oral data, it is also necessary to examine more written data as well. This is particularly important if the spoken corpora do in fact reflect a system that has been taken over by *tener que*, as the results from the present study suggest. Collecting additional written data from the 19<sup>th</sup> to the 21<sup>st</sup> centuries would further substantiate tendencies found in the present study and allow for comparisons across dialectal regions in order to account for and trace these changes.

Besides accounting for changes in the system, examining spoken and written data from more recent centuries also allow for the examination of sociolinguistic factors. Although age and sex did not show any differences in terms of the choice of obligation marker in the present study, research on other obligation markers have shown differences in use across generations, sex and education level (e.g. Tagliamonte and

Smith, 2006; Blas Arroyo, 2010). The inclusion of other social factors, such as education level, may provide a clearer picture of why and how some of the changes in the system occurred. Including more written and spoken data would also likely provide enough data to apply the comparative method (Tagliamonte & Poplack, 1988; Tagliamonte, 2006).

Taking into account the limited number of occurrences of *tener que*, *tener de* and *haber que* in the earlier centuries, it was impossible to incorporate the comparative method as a measure of change. This is in part because *haber de* accounted for more than 88% of all occurrences during the 16<sup>th</sup> century. A larger corpus, however, may contain a sufficient number of examples of the other three obligation markers to apply the comparative method. Recall that the adoption of the comparative method into variationist studies is a valuable tool for determining the source of change, establishing the status of grammaticalizing forms, as well providing a way to look at variation with a more fine-grained perspective. This is done by comparing cross-dialectal varieties or different time periods and identifying areas in the grammar that may be possible sources of change. The discovery of these sources can elucidate possible directions of change at a micro-level and corroborate assertions on the macro-level. In this case comparative analyses could further help in substantiating universal claims of paths of development as well as provide new information on how these changes take effect.

A future study should consider including the obligation marker *deber (de)* as a possible variant. Although this variant does not express strong obligation, which was the focus of the present study, incorporating it into the analysis may shed some light on changes that are occurring in the system as these markers move towards epistemicity.

This suggestion is based primarily on results from qualitative studies on obligation systems in other languages. For example, Myhill (1995: 160) found that the English obligation system underwent a general shift from using modals that were more focused on speaker control and societal norms, like we see in the example 5.1, to newer markers that focus less on power and more on 'interactive factors such as mutual cooperation, emotional appeals, advice, apologies or threats', like we see in example 5.2. Although he recognizes that there are exceptions in his data, he believes that this is a general trend in the history of obligation markers.

5.1 my daughter **shall** marry you (Myhill, 1995: 160)

5.2 'you've **got to** help me (emotional plea) (Myhill, 1995: 160)

Smith (2003: 263) seems to echo Myhill's claims in her explanation on the gradual disappearance of English 'must'. She posits that because it is prototypically subjective and insistent, sometimes authoritarian-sounding, root MUST is likely to be increasingly avoided in a culture where overt markers of power or hierarchy are much less in favour...'. In order to consider whether or not a similar trend is occurring in Spanish, an extensive examination of all obligation markers, including those that have been identified as being related to social norms in the literature (e.g. *deber (de)* in Molina Plaza, 2005) is required.

Traugott (1989) tries to connect changes in deontic modals to the social motivations behind obligation. She claims that any obligation that is based in morals, reason, law or divine decrees is enforced by some outside force and therefore cannot be considered subjective (1989: 39), contrary to what Smith (2003) is claiming. If we assume that obligation based in morals, reason etc. are more authoritarian, due to the

outside force, then we may be able to relate authoritarian with less grammaticalized obligation markers. If that is indeed the case, we can deduce that the movement from agent-oriented deonticity to subjective, epistemic meanings is also paralleled by a social shift from more authoritarian to less authoritarian (or what Myhill calls more interactive obligation, 1995). Nevertheless, a more inclusive study would be useful to corroborate this.

In sum, any further examination of this system over a time period of more than a century will provide a more reliable overview of the obligation system.

## **5.2 General Summary**

This study looked at the development of obligation in Mexican Spanish from the 16<sup>th</sup> through the 21<sup>st</sup> century. Considering the dearth of variationist research available on Spanish modals this study was an important step in determining whether the obligation system in Spanish is undergoing a change. There were several approaches adopted in an effort to reveal both the path of change of each of these markers, as well as to corroborate (or not) the postulations of universal paths of change in deontic modality.

As a result of the narrow scope that has been utilized thus far, little evidence has surfaced on how obligation markers are used in Spanish in both text and oral speech. This study sought to fill this gap of knowledge by looking at the development of modality in Spanish both diachronically and synchronically by adopting a variationist and comparative sociolinguistic approach (Poplack and Meechan, 1998; Poplack and Tagliamonte, 2001; Tagliamonte, 2002; Tagliamonte and Smith, 2006). In addition, an examination of the relative frequencies of these four obligation markers across time in a variety of contexts was also employed. The aim was to uncover the linguistic and

extralinguistic factors that condition the use of one expression of obligation over another while at the same time substantiating universal claims of change in modality.

The variationist approach was adopted in part because a fundamental element of grammaticalization is variation. The areas in grammar where variation is found are ideal for examining how change and thus grammaticalization occurs. In the present study the variation was particularly noticeable in the later centuries when *haber que* and *tener que* increased in frequency. Through variable rule analysis, the similarities and differences among these constructions were identified. We found that, contrary to the existing literature on obligation markers in Spanish, these constructions are used in distinct contexts and are conditioned by different factors.

In addition to the variationist method, the comparative method was also employed. Again, through a variable rule analysis the comparative sociolinguistic method is used as a way to examine similarities and differences in the language across time essentially offering a new detailed way to construct the development of the obligation system. As a result of the comparative analysis we saw that, between the 19<sup>th</sup> and 20<sup>th</sup> centuries, verb type and direct objects were areas in the grammar where changes were taking place, particularly with *haber de* and *tener que*.

This study was the first to look at obligation in Spanish from a both diachronic and variationist perspective. The results of this study not only illustrated that this system is in flux but it also contributed to the existing literature on the grammaticalization of modals across languages. In regards to these constructions, we see that each marker does indeed have its own semantic space and, although these forms grammaticalize and move towards epistemicity, they are in distinct parts of this path and thus fulfill different

semantic roles. Throughout the centuries, several factors have played a role in determining which marker will be used in the various contexts. The factors are also indicators of the movement of these obligation markers along the grammaticalization path. In spite of the fact that many of these factors have been mentioned in the previous literature, very few have been included in quantitative research on obligation makers (cf. Cornillie, 2007). I was able to incorporate the claims made in the literature and operationalize these claims into factors and successfully use these as measures of subjectification.

Similar to work done on obligation in English (e.g. Tagliamonte & Smith 2006; Tagliamonte & D'Arcy, 2007), this study offers evidence supporting the claims of a universal path of grammaticalization, specifically the path proposed by linguists such as Bybee et al. (1994), Hopper and Traugott (2003) among others. In particular, evidence was provided for processes such as semantic weakening, decategorialization and general movement into contexts associated with weak epistemicity and strong epistemicity. The data substantiates assertions that obligation systems undergo longitudinal renewal and that new markers come into the system because the older forms are no longer expressive enough (Tagliamonte & D'Arcy, 2007)

Previous literature on Spanish obligation markers rarely recognizes the semantic differences among these obligation markers or attributes the differences in meaning to psychological motives (e.g. De Maeseneer, 1998; Sirbu-Dumitrescu D, 1988). We now have a concrete example of the various factors that condition the use of each of these obligation markers, which is evidence against the idea that these forms essentially mean the same thing. It also illustrates that the factors that condition the use of each

marker are not entirely dependent on the psychological motives of the speaker. The results of this study also go hand in hand with other, similar studies on obligation (Tagliamonte & D'Arcy, 2007; Tagliamonte & Smith 2006), which all show similar patterns.

Through the adoption of comparative analyses during the 19<sup>th</sup> and 20<sup>th</sup> centuries, this study uncovered patterns of change in the system that illustrated the movement of constructions into and out of the semantic space of obligation. We can observe where in the grammar these specific changes are occurring through the use of the comparative method, which to my knowledge has not been done on obligation markers in Spanish before. This type of analysis has provided a baseline from which to build on for future research on obligation markers in both Spanish and other languages. It illustrates how both a diachronic and synchronic variationist examination of the obligation system can expose the path of obligation markers within Spanish, and perhaps across languages.

The importance of frequency was also clearly noted in this study. A construction may occur in contexts that could implicate possible movement into new contexts, but without an increase in frequency of use, this movement is unlikely. Recall that change like grammaticalization occurs in high frequency forms. We saw an example of this with the construction *tener de*, which appeared in contexts associated with weak epistemicity as early as the 16<sup>th</sup> century. Despite this apparent association, the frequency of this form decreased over time and, eventually, this form fell out of use. This example offers further support for the usage-based model where we see that usage and frequency are important factors in language change.

Finally, this study also predicted a possible path of grammaticalization for obligation markers in Spanish, where these constructions enter the language via declarative sentences containing communication or action verbs, typically in present tense or imperfect. Further grammaticalization is shown to be characterized by increasing occurrence with psychological and emotion verbs, habitual temporal markers and contexts containing negation. As the frequency in these new contexts rises, they begin to be associated with weak epistemicity. If the grammaticalization continues the construction begins to appear more often with stative verbs, and less often when negation is present. There is also an increase of occurrence with temporal future markers and inanimate subjects and objects. This is the stage that is identified as strong epistemicity by Traugott (1989).

What appears to be one of the last stages of grammaticalization of these obligation markers as they move into epistemic contexts is the occurrence in contexts with existential subjects and the presence of preceding clitics. We did not see examples of counter examples of the principle of unidirectionality. All these constructions appeared to go along this path in the same direction thus supporting claims for unidirectionality made by linguists such as Haspelmath (2004), Taeymans (2004) and Hopper and Traugott (2003). While not all these constructions have grammaticalized to this last stage, we see that *haber de* and *tener que* are going in that direction and, as they enter into epistemicity, it will likely open this space up for a new obligation marker to take.

APPENDIX A  
LIST OF TEXTS USED IN WRITTEN CORPORA

Work	Author	Word Count	Century	Corpus
Carta del Cabildo de la ciudad de Mérida al Rey Don Felipe	Anónimo	987	16	CORDE
Carta anual del P. Francisco Ramírez	Anónimo	4289	16	CORDE
Carta de Don Luis de Velasco, virey de la Nueva España, al Emperador Don Carlos, recordándole la indicación que, en su Real nombre, le hizo el Secretario Francisco de Eraso, al conferírsele aquel vireynato, de permutarlo, a los tres años de servicio.	Velasco, Luis de	825	16	CORDE
Carta de Fray Pedro de Gante al Emperador D. Carlos	Gante, Fray Pedro de	4777	16	CORDE
Carta de la ciudad de Michoacan	Anónimo	1488	16	CORDE
Carta de naturales de la provincia de Tlascalala al Rey Don Felipe II, suplicando les concediera exenciones, títulos y privilegios en remuneración de los servicios de sus antepasados al tiempo de la conquista.	Anónimo	2184	16	CORDE
Carta del Cabildo de la ciudad de Mérida al Rey Don Felipe II, haciendo patente la necesidad que tenia aquella tierra de defensa contra los franceses luteranos, recomendando la gobernacion de Don Luis Céspedes.	Anónimo	987	16	CORDE
Carta del clérigo Pedro de Logroño al Rey Don Felipe II	Logroño, Pedro de	1120	16	CORDE
Carta del doctor Diego Quixada, alcalde de Mérida de Yucatán, á S.M., dando cuenta de algunas medidas de buen gobierno y administracion de justicia, y consultando la adopcion de otras.	Quijada, Diego	4157	16	CORDE
Carta del licenciado Alonso Zuazo al padre Fray Luis de Figueroa, prior de la Mejorada.	Zuazo, Alonso	3768	16	CORDE
Carta del licenciado Francisco Ceynos	Ceynos, Francisco	2224	16	CORDE
Carta del licenciado Marcos de Aguilar	Aguilar, Marcos de	2628	16	CORDE
Carta del padre fray Jerónimo de Mendieta	Mendieta, Fray Jerónimo de	12,918	16	CORDE
Carta del virey de la Nueva España Don Martín Enríquez	Enríquez, Martín	2875	16	CORDE
Carta-relacion del arzobispo de Mexico D. Pedro de Moya y Contreras, remitiendo al Rey Don Felipe II reservados informes personales del clero de su diócesis.	Moya y Contreras, Pedro de	8472	16	CORDE
Contrato adicional entre Juan Cromberger y Juan Pablos	Anónimo	591	16	CORDE
Demanda de Ceballos en nombre de Pánfilo de Narváez	Ceballos, Hernando de	3013	16	CORDE
Desposorio espiritual entre el pastor Pedro y la Iglesia mexicana	Pérez Ramírez, Juan	3875	16	CORDE
Fragmento de la visita hecha a don Antonio de	Anónimo	27,143	16	CORDE

Mendoza				
Historia verdadera de la conquista de la Nueva España	Díaz del Castillo, Bernal	138,000	16	website
Informe al rey por el cabildo eclesiástico de Guadalajara	Anónimo	9820	16	CORDE
Instrucción civil y militar a Francisco Cortés	Cortez, Hernán	2265	16	website
Los mayas de Yucatán	Landa, Diego de	20,087	16	CORDE
Memorial de Fr. Bartolomé de las Casas	Casas, Fray Bartolomé de las	839	16	CORDE
Memorial de fray Bartolomé de las Casas al Consejo de Indias	Casas, Fray Bartolomé de las	1421	16	CORDE
Origen de los mexicanos	Anónimo	9896	16	CORDE
Parecer de Alonso del Castillo	Castillo, Alonso del	462	16	CORDE
Parecer de Fray Domingo de Betanzos	Betanzos, Fray Domingo de	2997	16	CORDE
Probanza hecha en la Villa Segura de la Frontera	Anónimo	3632	16	CORDE
Real cédula dirigida a Juan Gutiérrez Calderón	Anónimo	393	16	CORDE
Relación hecha por el señor Andrés de Tapia, sobre la conquista de México.	Tapia, Andrés de	17,855	16	CORDE
Revelación sobre la reincidencia en sus idolatrías	Feria, Fray Pedro	4494	16	CORDE
Segunda carta del doctor Ceynos	Ceynos, Francisco	2696	16	CORDE
Tercera relación anónima de la jornada que hizo Nuño de Guzman	Anónimo	8986	16	CORDE
Apologética historia sumaria	Casas, Fray Bartolomé de las	577,295	16	CORDE
Capítulo de carta del rey al virrey de Nueva España	Anónimo	273	16	CORDE
Carta al Rey Don Felipe II, del virey de la Nueva España, Marqués de Villamanrique, dando cuenta del arribo del corsario inglés Francisco Drakc, al puerto de la Habana.	Marqués de Villamanrique	2468	16	CORDE
Carta annua del P. Francisco Ramírez	Anónimo	4289	16	CORDE
Carta de Diego de Ocaña	Ocaña, Diego de	5745	16	CORDE
Carta de Don Luis de Velasco, virey de la Nueva España, al Emperador Don Carlos,, recordándole la indicacion que, en su Real nombre, le hizo el Secretario Francisco de Eraso, al conferírsele aquel vireynato, de permutarlo, a los tres años de servicio.	Velasco, Luis de	825	16	CORDE
Carta de Fray Angel de Valencia, custodio, y otros religiosos de la órden de San Francisco, proponiendo los medios necesarios para doctrinar los indios del Nuevo Reino de Galicia y de la provincia de Mechoacan.	Valencia, Fray Ángel de	6391	16	CORDE
Carta de Fray Domingo de Betanzos	Betanzos, Fray Domingo de	1258	16	CORDE
Carta de Fray Jacobo de Tastera, y de otros religiosos de la órden de San Francisco, al Emperador D. Carlos, dándole cuenta del estado de sus misiones y de la buena disposición de los indios.	Tastera, Fray Jacobo de	1879	16	CORDE
Carta de Fray Martin de Valencia, y otros misioneros al emperador.	Valencia, Fray Martín de	1099	16	CORDE
Carta de Fray Miguel Navarro, y otros religiosos de	Navarro, Fray	781	16	CORDE

la orden de San Francisco, al Real Consejo de las Yndias recomendando al licenciado Muñoz.	Miguel			
Carta de Jerónimo López al emperador.	López, Jerónimo	5556	16	CORDE
Carta del arzobispo de Mexico D. Pedro de Moya y Contreras al presidente de los Reales consejos de Indias y Hacienda, sobre los conflictos ocasionados por la representacion de un entrmés, y otros sucesos.	Moya y Contreras, Pedro	7839	16	CORDE
Carta del contador Rodrigo de Albornoz	Albornoz, Rodrigo de	12,400	16	CORDE
Carta del ejército de Cortés al emperador	Anónimo	3297	16	CORDE
Carta del obispo de Mexico, Fray Juan de Zumarraga á Juan de Samano, secretario de S.M., haciéndole presente algunas necesidades de sus diocesanos, y rogándole que apoyara su proyecto de edificacion de colegios y monasterios para jóvenes de ambos sexos.	Zumárraga, Fray Juan de	4399	16	CORDE
Carta del obispo de Yucatán Fray Francisco de Toral, al adelantado de la Florida, Pedro Menéndez de Avilés, acusando el recibo de otra suya, anunciándole el pronto envío de bastimentos y dándole consejos para su buen gobierno.	Toral, Fray Francisco de	1275	16	CORDE
Carta del padre fray Jerónimo de Mendieta	Mendieta, Fray Jerónimo de	12,918	16	CORDE
Carta del virey de la Nueva España	Conde de Coruña	3806	16	CORDE
Carta del virey de la Nueva España	Enríquez, Martín	3977	16	CORDE
Carta inédita de Hernán Cortés	Cortez, Hernan	5961	16	CORDE
Cartas de relaciones	Hernán Cortez	246,000	16	website
Contrato celebrado entre Juan Cromberger	Anónimo	702	16	CORDE
Contrato de compañía celebrado entre Juan Cromberger y Juan	Anónimo	2298	16	CORDE
Crónica de la Nueva España	Cervantes de Salazar, Francisco	400,000	16	website
Cuarta relación anónima de la jornada que hizo Nuño de Guzmá	Anónimo	9177	16	CORDE
Daños e inconvenientes que se derivan, en lo temporal y en lo espiritual, de la ignorancia de los idiomas aborígenes.	Molina, fray Alonso de	760	16	CORDE
Expediente tramitado en León de Nicaragua	Anónimo	2403	16	CORDE
Genealogía y linaje de los Señores de Nueva España	Anónimo	6873	16	CORDE
Historia de los Indios de la Nueva España	Fray Toribio de Benavente	97,329	16	CORDE
Historia de los mexicanos por sus pinturas	Anónimo	13,001	16	CORDE
Historia general de las cosas de Nueva España	Sahagún, Fray Bernardino de	368,588	16	CORDE
Historia verdadera de la conquista de la Nueva España	Díaz del Castillo, Bernal	48,000	16	CORDE
Informe al rey por el cabildo eclesiástico de Guadalajara, acerca de las cosas de aquel reino.	Anónimo	9820	16	CORDE
Instrucción civil y militar a Francisco Cortés, para la expedición de la costa de Colima.	Cortés	2265	16	CORDE
Instrucciones al virrey de la Nueva España para que se divulgue el conocimiento de la lengua española	Anónimo	328	16	CORDE

entre los indios, a fin de evitar errores en la cristianización; tratándose asunto tan importante con la audiencia, obispo y misioneros.				
Leyes de los indios de Nueva España, Anáhuac o México.	Alcobiz, Fray Andrés de	2222	16	CORDE
Leyes y ordenanzas	Anónimo	9098	16	CORDE
Lo que pasó con Cristóbal de Tapia acerca de no admitirle por gobernador, con los procuradores de México y demás poblaciones, y los de Hernán Cortés.	Anónimo	4633	16	CORDE
Memoria de lo acaecido en esta ciudad después que el gobernador Hernando Cortés salió della, que fue a los doce días del mes de octubre de mil e quinientos e veinte e cinco años.	Anónimo	5228	16	CORDE
Memorial de don Alonso de Zurita	Zurita, Alonso de	4232	16	CORDE
Memorial del obispo Fray Bartolomé de las Casas y Fray Domingo de Santo Tomás.	Casas, Fray Bartolomé de las	2244	16	CORDE
Ordenaciones generales dadas por el padre Diego de Avellaneda por las que debían regirse las provincias mexicanas de la Compañía de Jesús.	Anónimo	366	16	CORDE
Ordenanzas militares y civiles mandadas pregonar por don Hernando Cortés en Tlaxcala, al tiempo de partirse para poner cerco a México.	Anónimo	2435	16	CORDE
Parecer de don Sebastián Ramírez de Fuenleal, obispo de Santo Domingo, y presidente de la real audiencia de Nueva España.	Ramírez de Fuenleal, Sebastián	8928	16	CORDE
Primera parte de los problemas y secretos maravillosos de las Indias	Cárdenas, Juan de	88,739	16	CORDE
Primera relación anónima de la jornada que hizo Nuño de Guzmán a la Nueva Galicia	Anónimo	3312	16	CORDE
Proceso de Pedro de Ocharte	Anónimo	22,780	16	CORDE
Real Cédula a los Oficiales de la Casa de la Contratación de Sevilla.	Anónimo	313	16	CORDE
Real Cédula al arzobispo de México pidiendo informes sobre las características de la tierra.	Anónimo	1049	16	CORDE
Real Cédula dirigida a Juan Cromberger, por la que se le concede privilegio para la impresión de la Doctrina en lengua de indios de Mechoacán.	Anónimo	448	16	CORDE
Real Cédula y cuestionario para la formación de descripciones geográficas eclesiásticas del arzobispado de México.	Anónimo	1411	16	CORDE
Relación de la conquista de los Teules chichimecas que dio Juan de Sámano.	Sámano, Juan de	11,332	16	CORDE
Relación de la entrada de Nuño de Guzmán, que dio García del Pilar, su intérprete.	Pilar, García del	4949	16	CORDE
Segunda relación anónima de la jornada que hizo Nuño de Guzmán	Anónimo	4331	16	CORDE
Historia eclesiástica indiana	Mendieta, Fray Jerónimo	334,813	17	CORDE
Denuncia del indio tarasco Antonio Joan Apatzi por haber sido solicitado por un fraile franciscano.	Anónimo	1085	17	CORDE
Fragmentos de una historia de la Nueva Galicia	Tello	39,077	17	CORDE
Historia de la Nueva México	Villagrà, Gaspar	78,687	17	CORDE
Historia eclesiástica indiana	Mendieta, Fray Jerónimo	334,813	17	CORDE
Infortunios de Alonso Ramírez	Sigüenza y	17,793	17	CORDE

Ordenanzas de los maestros de escuelas de la ciudad de México	Góngora Anónimo	1182	17	CORDE
Los sirgueros de la Virgen sin original pecado	Bramón	18,585	17	CORDE
El pastor de nochebuena	Palafox y Mendoza	33,489	17	CORDE
Fragmento de una Real Cédula	Anónimo	806	17	CORDE
Historia de la nación chichimeca	Ixtlilxochitl	100,502	17	CORDE
Carta autógrafa de Juan Bruno Eusebio de Palma, tirador de oro y plata, denunciando a un solicitante	Palma, Juan Bruno Eusebio de	487	18	CORDE
Historia Antigua de México	Clavijero, Francisco Javier	334,302	18	CORDE
La portentosa vida de la muerte	Bolaños, Fray Joaquín	65,525	18	CORDE
Memoria sobre la construcción de sumideros para purificar la atmósfera	Basadre, Vicente	3026	18	CORDE
El Zarco, episodio de la vida mexicana en 1861-63	Altamirano, Ignacio Manuel	51,187	19	CORDE
Gil Gómez, el insurgente novela histórica	Díaz Covarrubias, Juan	81,934	19	CORDE
La Quijotita y su prima	Fernández de Lizardi	172,696	19	CORDE
Memoria sobre la necesidad y utilidades de la construcción de un camino carretero desde Veracruz a México.	Austria, José Donato de	7008	19	CORDE
Clemencia	Altamirano, Ignacio Manuel	51,118	19	CORDE
Cuentos del General	Riva Palacio	31,086	19	CORDE
Historia de Chucho el Ninfo	Facundo (José Tomás de Cuéllar)	67,761	19	CORDE
La bola	Rabasa, Emilio	38,889	19	CORDE
La gran ciencia	Rabasa, Emilio	39,827	19	CORDE
La Linterna Mágica	José Tomás de Cuéllar	35,900	19	internet archive
Noches tristes y día alegre	Fernández de Lizardi	25,726	19	CORDE
Oración patriótica en la primera fiesta conmemorativa de la MÉXICO	Barquera, Juan Wenceslao	3570	19	CORDE
Suprema Ley	Gamboa, Federico	114,215	19	CORDE
Tomóchic	Frías Heriberto	68,300	19	internet archive
Cosas de cualquier familia	Medina	52,310	20	CREA
Dalia	Martin del Campo	35,420	20	CORDE
Duerme	Boullosa	30,215	20	CREA
El apando	Revueltas	10,570	20	CORDE
El canto de la grilla	Rubín	47,527	20	CORDE
El error de la luna	Aguilar Camín, Héctor	61368	20	CORDE

Imposibilidad de los cuervos	Padilla	28,676	20	CREA
La casa de las mil vírgenes	Azuela, A	98,104	20	CREA
La frontera increíble	Revueltas	1961	20	CORDE
La muerte de Artemio Cruz	Fuentes, Carlos	95,060	20	CORDE
La región más transparente	Fuentes, Carlos	145,067	20	CORDE
Por vivir en quinto patio	Alatriste	62,798	20	CREA
Sinfonía pastoral	Revueltas	8159	20	CORDE
¿Enfermera, doctora o santa?	Olivera Figueroa	75,376	20	CREA
Andanzas del indio Vicente Alonso	Montaño Hurtado	24,799	20	CREA
Arráncame la vida	Ángeles	75,001	20	CREA
Como agua para chocolate	Esquivel, L	53,874	20	CREA
Cristóbal Nonato	Fuentes	196,455	20	CORDE
De Los Altos	Chao Ebergenyi	149,939	20	CREA
Dios en la tierra	Revueltas	2126	20	CORDE
Dos crímenes	Ibargüengoitia, Jorge	53,934	20	CREA
El batallador	Chávez Jr., Gilberto	136,959	20	CREA
El callado dolor de los tzotziles	Rubín	34,751	20	CORDE
El error de la luna	Aguilar Camín, Héctor	61,368	20	CREA
El lenguaje de nadie	Revueltas	3358	20	CORDE
El perro de la escribana o Las Piedecabras	Mendoza	24,644	20	CREA
El tamaño del infierno	Azuela, A	141,269	20	CORDE
Fábrica de conciencias descompuestas	Gerardo	40,924	20	CREA
Hegel y yo...	Revueltas	3859	20	CORDE
Jesús el bisabuelo y otros relatos	Espinosa	43,129	20	CREA
Juegos florales	Pitol	59,357	20	CREA
La caricia rota	Aguilera, Nuri	32,136	20	CREA
La luciérnaga	Azuela, Mariano	39,624	20	CORDE
Los de abajo	Azuela, Mariano	34,644	20	CORDE
Morir en el Golfo	Aguilar Camín, Héctor	83,261	20	CREA
Palinuro de México	Paso	276,537	20	CREA
Que la carne es hierba	Campos, Marco	24,270	20	CREA
Quién como Dios	González	136,658	20	CREA
Setenta veces siete	Elizondo	75,357	20	CREA
Un grito desesperado. Novela de superación personal	Cauhtémoc	49,588	20	CREA
Una piñata llena de memoria	Leyva	101,865	20	CREA

APPENDIX B  
SAMPLE QUESTIONS FOR SOCIOLINGUISTIC INTERVIEW

1. Demografía:

¿Cómo se llama?

¿Cuál es su dirección?

¿Hace cuánto tiempo que vive allí?

¿Dónde nació?

¿Ha vivido en otros lugares? ¿Dónde?

¿Dónde nacieron y crecieron sus padres? ¿sus abuelos? ¿su marido/esposa?

¿Trabaja? ¿Dónde?

¿A qué se dedican sus padres? ¿su marido/esposa?

¿Cuántos años de escuela pudo terminar?

¿Cuál fue el primer trabajo que tuvo al salir de la escuela?

Me puede contar algo sobre su casa/apartamento?

¿Qué tipo de casa es?

¿Habla otras lenguas además del español?

2. El barrio

Este barrio parece interesante. ¿Hace cuanto tiempo que vive aquí? ¿Qué tipo de gente vive en esta calle? ¿en esta área?

¿Por qué se mudaron aquí sus padres? (o usted)

¿Queda lejos del trabajo de su padre/suyo?

¿Cómo ha cambiado el barrio en su vida?

¿Se siente que el barrio sea tan seguro como era cuando era niño/a? ¿por qué o por qué no?

¿Es un barrio donde la gente se habla?

¿Conoce Ud. a sus vecinos? ¿Cómo son?

Se dice que hoy en día todos están demasiado ocupados para pasar por las casas de los vecinos para charlar, ¿qué cree Ud.?

¿Por qué cree que (no) ha cambiado?

¿Hay que gente aquí que conoce tan bien que no le necesita anunciar su visita?

¿Hay que gente de aquí que pasa por su casa para charlar?

¿Hay que un lugar en esta vecindad donde la gente se reúne?

¿Dónde se puede ir para un refresco o un té/café por la tarde?

¿Hay que un bar que frecuenta?

¿Dónde se reúne la gente fuera de casa?

¿Hay que gente con quienes quiere pasar más tiempo, pero no puede?

¿Por qué no la ve tanto ya?

¿Cree que la comunidad/barrio puede ser más unida/o? ¿Cómo?

¿Cuál es la cosa que le gusta más del barrio? ¿Cuáles cosas le hace sentir bien/mal de su barrio?

3. Costumbres sociales

¿Hay que gente por aquí que no son familiares y con quienes pasa mucho tiempo?

¿Viven cerca? ¿Por dónde? ¿Qué hacen cuando se reúnen?

¿Sale con amigos? ¿Qué hacen?

¿Va a bailes?

¿Qué tipo de entretenimiento *hay que* por aquí?

¿Qué tipo de baile hace?

¿Cómo se viste la gente cuando va?

#### 4. Familia

¿Sabe hace cuánto tiempo que vive su familia aquí en este pueblo/estado/país?

¿De dónde venían?

¿Recuerda cuando llegó a este pueblo/estado/país?

¿Cómo era cuando era niño/a?

¿Era travieso? ¿Qué tipo de cosas traviesas hacía?

¿Le castigaron? ¿Quiénes?

¿Se le culparon por cosas que no había hecho?

¿Sus padres imponían un toque de queda?

¿Qué pasó si no lo obedeció?

¿Sus padres ya tenían ideas de lo que querían que Ud. fuera cuando se crecía?

¿Tiene hermanos? ¿Cuántos?

¿Cómo era su relación con ellos?

¿Cómo era ser el mayor/menor/en el medio de los hermanos?

¿Cree que sus hermanos pudieron hacer cosas que sus padres no le permitieron hacer a Ud.? ¿Puede darme un ejemplo?

¿Pasaba mucho tiempo con sus \_\_\_\_\_?

abuelos, tíos y tías, primos etc.

¿por qué o por qué no?

¿Iban de vacaciones mucho?

¿Adónde? ¿Cómo llegaron?

¿Había una vez en que alguien en su familia le avergonzó por algo que había dicho o hecho?

¿Qué pasó?

¿Cómo reaccionó?

¿Qué puede decir ahora retrospectivamente?

#### 5. Trabajo

¿Cómo fue su primer trabajo?

¿Cuántos años tenía cuando empezó a trabajar?

¿Recuerda cuánto ganó?

¿Recuerda cómo quería gastar ese dinero?

¿A qué se dedicaban sus padres?

¿A qué se dedica Ud.?

¿Qué le gustaría hacer como trabajo?

#### 6. La comida

Se dice que mucha gente tiene su mejor comida el domingo, ¿era así en su familia?

¿Qué tipo de comida comían en su niñez?

Mucha gente recuerda ciertos platos que hacían sus madres/padres, ¿se acuerda de un plato especial que hacía su madre/padre?

¿Qué comía su \_\_\_\_\_? abuelo, abuela, padre, madre  
¿Cocina Ud.? ¿Coce? ¿Qué tipo de comida le gusta cocer/cocinar?  
¿La comida de hoy es diferente de la comida de su niñez? ¿Cómo? ¿Por qué?

#### 7. Medicina popular

¿Iba la gente mucho al doctor cuando era niño/a? ¿Iba Ud. cuando estaba enfermo/a?  
Se dice que la gente dependía mucho de la medicina y los remedios populares,  
¿recuerda ese tiempo?

¿Recuerda un remedio particular?

Me acuerdo que mi abuelo tomaba mucho \_\_\_\_\_ (ajo, miel, limón, etc.), ¿Ha escuchado de ese remedio?

¿Qué hace cuando se enferma? ¿Qué hace para evitar los resfriados o la gripe?

#### 8. Jóvenes/padres

Se dice que ser joven hoy en día es muy diferente que serlo cuando era joven Ud.,  
¿está de acuerdo?

¿Cuál es la diferencia? ¿Por qué?

Puede comparar que hacía para divertirse cuando era joven con las cosas que hacen los jóvenes de hoy para diversión.

Se dice que los padres de hoy son más estrictos, ¿está de acuerdo?

¿Qué no le gustaba de sus padres, hermanos?

¿Gastabas bromas a sus hermanos?

¿Qué fue la mejor broma que gastó? ¿La más chistosa? ¿La peor?

APPENDIX C  
COMPLETE LIST OF FACTOR GROUPS AND FACTORS

Verb Class

- Action (e.g. *buscar* 'look for', *caminar* 'walk')
- Psychological /emotional (e.g. *sorprender* 'surprise', *pensar* 'think', *entender* 'understand')
- Communication (*hablar* 'talk', *gritar* 'shout', *aclarar* 'to clarify')
- Perception (*ver* 'see', *oír* 'hear')
- Copular (*ser* 'be', *estar* 'be' tener cuidado 'be careful')
- The verbs *haber*/*tener*
- The verb *hacer*
- Other (*morir* 'die', *vivir* 'live')

Animacy/grammatical person

- 1<sup>st</sup> person animate (e.g. *tenemos que ponemos al día*. 'we have get updated'(TQ))
- 2<sup>nd</sup> person animate (e.g. *has de morir en Guayaquil* 'you must die in Guayaquil' (HD))
- 3<sup>rd</sup> person animate (e.g. *ya ves que mi mamá siempre tiene que ganar*. 'You see that my mom always has to win'(TQ))
- 3<sup>rd</sup> person inanimate (*mi angustia hubo de alcanzar su punto más alto* 'my anguish must have reached its maximum point' (HD))
- other (entities that have been given animate characteristics for example: *Puebla tendría que darle al Presidente la recepción* 'Puebla (the city) must give the President the reception (TQ); Mastretta, CREA)

Type of Sentence

- Interrogative (e.g. *¿Tuviste que pasar por allí, por Colima?* 'You had to pass through Colima?' (TQ))
- Declarative (*Hay que dar el apoyo a él*. 'We have to give our support to him'(HQ))

Polarity

- Negative (*Yo no tenía que pasar la noche fuera de mi casa*. 'I didn't have to spend the night away from my house'(TQ))
- Positive (*Teníamos que ir a una plática*. 'We had to go to the talk'(TQ))

Direct Objects

- Animate (*tenía que pagar el rico*. 'I had to pay the rich man'(TQ))
- Inanimate (*tiene que pagar uno renta*. 'one has to pay rent'(TQ))
- Oblique (*tengo que ir con una persona que me hable*. 'I have to go with someone who talks to me'(TQ))
- Middle Voice (e.g. *y si te cuesta tanto se tiene que traer*. 'and if it costs you a lot you have to bring it yourself'(TQ))
- Other

Indirect Objects

- Present (*tenía que dar un respecto a nuestros padres*. 'I had to give respect to our parents')
- Not present (*tienes que ganar mi confianza* 'you have to earn my confidence')

Preceding clitic

Present (*Usted me tiene que enseñar*. 'you (formal) have to teach me'(TQ))  
Not present (*teníamos que salir más temprano* 'we used to have to leave earlier'(TQ))

Tense-aspect-mood

For example,

present: (*tiene que aceptar* 'he has to accept' (TQ))

imperfect: (*no había que irnos a veces* 'we didn't have to go sometimes'(HQ))

imperfect subjunctive: (*sí pasara el examen voy a tener que salir* 'if I pass the exam, I'll have to graduate (TQ))

Temporal adverbs

For example,

specific: (*ahora ya tengo que cerrar los ojos* 'now I have to close my eyes'(TQ);

Fuentes, CORDE)

non-specific: (*quizá muy pronto tengamos que salir del país* 'maybe, very soon we'll have to leave the country'(TQ); Azuela, CORDE)

other: (*y tuvo que pensar* 'and he had to think'(TQ); Medina, CREA)

Lexical verb type

For example:

*seguir*: (*Hay que seguir adelante* 'One must continue on'(HQ))

*tener*: (*ya has de tener dinerito* 'one must already have money' (HD))

Sex

Male

Female

Age

20-45 years

greater than 45 years

## LIST OF REFERENCES

- AARON, JESSI ELANA. 2004. The gendered use of *salirse* in Mexican Spanish: Si me salía yo con las amigas, se enojaba. *Language in Society* 33(5).585-607.
- AARON, JESSI ELANA. 2006. Me voy a tener que ir yendo: A Corpus-Based Study of the Grammaticization of the *ir a + INF* Construction in Spanish. *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, ed. by Nuria Sagarra and Almeida Jacqueline Toribio, 263-72. Somerville, MA: Cascadilla Proceedings Project.
- AARON, JESSICA ELANA. 2007. *Variation and change in spanish future temporal expression: Rates, constraints, and grammaticization*. Albuquerque, NM: UNM dissertation.
- AARON, JESSI ELANA. 2010. Pushing the envelope: Looking beyond the variable context. *Language Variation and Change* 22.1-36.
- AARON, JESSI. ELANA. AND RENA TORRES CACOULLOS. 2005. Quantitative measures of subjectification: A variationist study of Spanish *salirse*. *Cognitive Linguistics* 16.4.607-633.
- ALAMILLO, ASILLA REIG. 2009. Cross-dialectal variation in propositional anaphora: Null objects and propositional *lo* in Mexican and Peninsular Spanish. *Language Variation & Change*, 21(3).381-415
- AMMON, U. 2004. *Sociolinguistics: An international handbook of the science of language and society*. Berlin: Walter de Gruyter.
- ASH, SHARON. 2002. Social Class. *The handbook of language variation and change* ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 402-422. Malden, MA: Blackwell.
- AXELROD, MELISSA. 1999. Lexis, grammar, and grammatical change: The koyukon classifier prefixes. *Functionalism and formalism in linguistics: Volume ii: Case studies*, ed. By In Darnell, Michael, Moravcsik, Edith, Newmeyer, Frederick, Noonan, Michael, & K. Wheatley, 39-58. Amsterdam: John Benjamins.
- BAILEY, GUY. 2002. Real and Apparent Time. *The handbook of language variation and change* ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 312-332. Malden, MA: Blackwell.
- BARLOW, MICHAEL. 2000. Usage, Blends and Grammar. *Usage-based models of language*, ed. by Michael Barlow, and Suzanne Kemmer, 315-346. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information.

- BARLOW, MICHAEL AND SUZANNE KEMMER, 2000. *Usage-based models of language*. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information.
- BAYLEY, ROBERT. 2002. The quantitative paradigm. *The handbook of language variation and change* ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 117-141. Malden, MA: Blackwell.
- BIBER, DOUGLAS. 2000. Investigating Language Use through Corpus-Based Analyses of Association Patterns. *Usage-based models of language*, ed. by Michael Barlow, and Suzanne Kemmer, 287-314. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information.
- BLACKWELL. 1987. *Observing and analysing natural language: A critical account of sociolinguistic method*. Oxford, OX, UK: Blackwell.
- 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* 20.85-126.
- BLAS ARROYO, JOSÉ LUIS. 2010. Confluencia de normas sociolingüísticas en un hecho de variación sintáctica: Factores sociales en la selección de la variante "deber de" + infinitivo (vs. "deber") en un corpus de español peninsular. *Hispania* 93(4).624-649
- BOLINGER, DWIGHT. 1977. *Meaning and form*. London: Longman.
- BRINTON, LAUREL J. 1991. The Origin and Development of Quasimodal 'have to' in English. *The Workshop on Verbal Periphrases*, Amsterdam. Online: <http://faculty.arts.ubc.ca/lbrinton/>
- BRINTON AND TRAUOGOTT. 2003. *Lexicalization and language change*. Cambridge, England: Cambridge University Press.
- BRINTON AND TRAUOGOTT. 2005. Lexicalization and grammaticalization all over again. *Historical Linguistics 2005: Selected papers from the 17<sup>th</sup> international conference on historical linguistics*, ed. by Joseph C. Salmons and Shannon Dubenion-Smith, 3-31. Amsterdam Netherlands: John Benjamins Publishing Co.
- BYBEE, JOAN L. 1987. The semantic development of past tense modals in English and other languages. *Modality in grammar and discourse*, ed. by Joan Bybee and Suzanne Fleischman, 503-517. Amsterdam, Netherlands: John Benjamins Publishing Co.
- BYBEE, JOAN L. 1998. A functionalist approach to grammar and its evolution. *Evolution of Communication* 22.249-278.
- BYBEE, JOAN L. 2001. *Phonology and language use*. Cambridge, England: Cambridge University Press.

- BYBEE, JOAN. 2003a. Mechanisms of change in grammaticization: The role of frequency. *The handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 602–23. Oxford: Blackwell
- BYBEE, JOAN L. 2003b. Cognitive processes in grammaticalization. *The new psychology of language: Cognitive and functional approaches to language structure*, ed. by Michael Tomasello, 2.145-167. Mahwah, NJ: Lawrence Erlbaum Associates.
- BYBEE, JOAN L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711-733.
- BYBEE, JOAN. L., AND PAGLIUCA, WILLIAM. 1985. Crosslinguistic comparison and the development of grammatical meaning. *Historical semantics: Historical word-formation* ed. by Jacek Fisiak, 59–83. Berlin: Mouton de Gruyter.
- BYBEE, JOAN. L., REVERE D. PERKINS, and WILLIAM PAGLIUCA. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- BYBEE, JOAN & SANDRA THOMPSON. 2000. Three frequency effects in syntax. *Berkeley Linguistics Society* 23.65-85.
- BYBEE, JOAN, AND SUZANNE FLEISCHMAN. 1995. Modality in grammar and discourse: An introductory essay. In *Modality in grammar and discourse.*, ed. by Joan Bybee, Suzanne Fleischman, 1-14. Amsterdam, The Netherlands: John Benjamins Publishing Co.
- CAMPBELL, LYLE. 1998. *Historical linguistics: an introduction*. Edinburgh: Edinburgh University Press.
- CASANOVA SEUMA, LOURDES. 1999. El sujeto en catalán coloquial. *Revista Española de Lingüística* 29(1).105-131.
- CEDERGREN, HENRIETTA J., and DAVID SANKOFF. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50.333-355.
- CHOI, JINNY. 1998. *Languages in contact: a morphosyntactic analysis of Paraguayan Spanish from a historical and sociolinguistic perspective*. Washington, D.C.: Georgetown University dissertation
- COATES, JENNIFER. 1983. *The Semantics of the modal auxiliaries*. London: Croom Helm.
- COMPANY COMPANY, CONCEPCION. 1994. *documentos lingüísticos de la Nueva España: Altiplano central*. México: UNAM
- COMPANY COMPANY, CONCEPCION. 2004. Grammaticalization or degrammaticalization? the reanalysis and subjectivization of verbs as discourse markers in the history of spanish. *Revista de Filología Española* 84(1).29-66.

- COMPANY COMPANY, CONCEPCION. 2006. Subjectification of verbs into discourse markers: Semantic-pragmatic change only? *Belgian Journal of Linguistics* 20.97-121.
- CORNILLIE, BERT. 2007. *Evidentiality and epistemic modality in Spanish (semi)auxiliaries. A cognitive-functional approach*, Berlin: Mouton de Gruyter.
- CORNILLIE, BERT, DE MULDER, WALTER, VAN HECKE, TINE AND DIETER VERMANDERE. 2009. Modals in Romance Languages, *Modals in the Languages of Europe. A reference work*, ed. by Björn Hansen., Ferdinand de Haan Berlin, 511-561. Berlin: Mouton de Gruyter.
- DAVIES, MARK. 1995. Analyzing syntactic variation with computer-based corpora: the case of modern Spanish clitic climbing. *Hispania* 78(2). 370-380.
- DE MAESENEER, RITA. (1998). On some problems related to periphrases of obligation. [Sobre algunos problemas relacionados con las perífrasis obligativas] *Linguística Antverpiensia* 32.39-53.
- DÍAZ-CAMPOS, MANUEL. 2011. *The handbook of Hispanic sociolinguistics*. Malden, MA: Wiley-Blackwell.
- ECKART, PENELOPE, 1997. Age as a Sociolinguistic Variable. In *The handbook of sociolinguistics* ed. by Florian Coulmas, 151-168. Oxford, UK: Blackwell Publishers.
- ERMAN, BRITT, AND BEATRICE WARREN. 2000. The idiom principle and the open choice principle. *Text* 20.29-62.
- FENTE GÓMEZ, RAFAEL, JESÚS FERNÁNDEZ ALVAREZ, AND LOPE G. FEIJÓO. 1983. *Perífrasis verbales*. Madrid: EDI-6
- FERNANDEZ DE CASTRO, FELIX. 1999. *Las perífrasis verbales en el español actual*. Madrid: Gredos.
- FINEGAN, EDWARD. 1995. Subjectivity and subjectivisation: an introduction. *Subjectivity and subectivisation: Linguistic perspectives*, ed, by D.Stein and S. Wright, 1-15. Cambridge: Cambridge University Press.
- FLEISCHMAN, SUZANNE. 1982. *The future in thought and language: Diachronic evidence from Romance*. Cambridge, England: Cambridge University Press.
- GIVON, TALMY. 1984. *Syntax. A functional-typological introduction*, Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Co.
- GÓMEZ TORREGO, LEONARDO. 1988. *Perífrasis verbales: Sintaxis, semántica y estilística*. Madrid: Arco/Libros.

- GUTIÉRREZ , MANUEL J. 2010 La media subida de clíticos en el español de Houston. *Nueva Revista de Filología Hispánica* 58(2).645-662
- GUY, GREGORY. 1988. Advanced VARBRUL analysis. *Linguistic Contact and Change: Proceedings of the Sixteenth Annual Conference on New Ways of Analyzing Variation*, ed. by Kathleen Ferrara, Becky Brown, Keith Walters and John Baugh, 124-36. Austin, TX: University of Texas Department of Linguistics.
- HAIMAN, JOHN (1994). Ritualization and the development of language. *Perspectives on grammaticalization*, ed. by William Pagliuca, 3-28. Amsterdam: John Benjamins Publishing Co.
- HASPELMATH, MARTIN. 2004. Does linguistic explanation presuppose linguistic description? *Studies in Language* 28(3).554-579 (special issue on linguistic evidence, guest edited by Martina Penke and Anette Rosenbach).
- HEINE, BERND. 1995. Agent-Oriented vs. Epistemic Modality: Some Observations on German Modals. *Modality in grammar and discourse* ed. by Joan L. Bybee and Suzanne Fleischman, 17-55. Amsterdam: John Benjamins.
- HEINE, BERND, ULRIKE CLAUDI, AND FRIEDERIKE HÜNNEMEYER. 1991. Grammaticalization: A conceptual framework. Chicago: University of Chicago Press.
- HENGEVELD KEES. 1988. Illocution, Mood and modality in a functional grammar of Spanish. *Journal of Semantics* 6.227-269
- HOPPER, PAUL J., AND ELIZABETH CLOSS TRAUOGOTT. 1993. *Grammaticalization. Cambridge textbooks in linguistics*. Cambridge, England: Cambridge University Press.
- HOWE, CHAD & SCHWENTER, SCOTT A. 2008. Variable constraints on past reference in dialects of Spanish. *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*, ed. by Maurice Westmoreland & Juan Antonio Thomas, 100-108. Somerville, MA: Cascadilla Proceedings Project.
- HYMAN, LARRY: 1984. Form and Substance in Language Universals. *Explanations for Language Universals*, ed. Brian Butterworth, Bernard Comrie, and Östen Dahl. Berlin: Mouton. 67-85.
- INTERNET ARCHIVE (2002a). About the Internet Archive. Online: <http://www.archive.org/about/about.php>.
- KASTOVSKY, DIETER. 1991. *Historical English syntax*. Berlin: Mouton de Gruyter.
- KIEFER, FERENC. 1987. On defining modality. *Folia Linguistica* XXI. 67–94.

- KIM, MINJU. 2009. The intersection of the perfective and imperfective domains: A corpus-based study of the grammaticalization of Korean aspectual markers. *Studies in Language*, 33(1), 175-214.
- KLEIN-ANDREU, FLORA. 1991. Losing ground: A discourse-pragmatic solution to the history of -ra in Spanish. *Discourse-pragmatics and the verb: the evidence from Romance*, ed. by Suzanne Fleischman & Linda R. Waugh, 164-78. London: Routledge.
- KRUG, MANFRED. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin, New York: Mouton de Gruyter.
- KURYLOWICZ, JERZY. 1976[1965]. The evolution of grammatical categories. repr. in Jerzy Kurylowicz, *Esquisses linguistiques*, Vol. 2, 38-54. München: Fink.
- LABOV, WILLIAM. 1966. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- LABOV, WILLIAM. 1972. *Sociolinguistic Patterns*. Philadelphia: The University of Pennsylvania Press.
- LABOV, WILLIAM. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford, UK: Blackwell.
- LABOV, WILLIAM. 2008. Quantitative reasoning in linguistics.  
<http://www.ling.upenn.edu/~wlabov/Papers/QRL.pdf>
- LANGACKER, RONALD. 2000. A Dynamic Usage-Based Model. *Usage-based models of language*, ed. by Michael Barlow, and Suzanne Kemmer, 1-64. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information.
- LAVENDERA, B. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7.171-182.
- LEHMANN, W. P., & MALKIEL, YAKOV. 1982. *Perspectives on historical linguistics*. Amsterdam, Netherlands: John Benjamins Publishing Co.
- LEVEY, STEPHEN. 2012. Understanding children's non-standard spoken English: a perspective from variationist sociolinguistics. *Language and Education* 1-17.
- LUNA TRAILL, ELIZABETH; 1980: *Sintaxis de los verboides en el habla culta de la ciudad de México*. México: UNAM.
- LUNN, PATRICIA. 1995. Modality in grammar and discourse: An introductory essay. In *Modality in grammar and discourse.*, ed. by Joan Bybee, Suzanne Fleischman, 1-14. Amsterdam, Netherlands: John Benjamins Publishing Co.
- LYONS, JOHN. 1977. *Semantics*. Cambridge: Cambridge University Press.

- MARRANO, ANN MARIE. 1997. *The syntax of modality: A comparative study of epistemic and root modal verbs in Spanish and English*. Georgetown, WA: Georgetown University dissertation.
- MARTINEZ-DIAZ, EVA. 2003. La frecuencia de use de “haber” y “tener” en las estructuras perifrásticas de obligación. Algún fenómeno de variación en el español de Cataluña. *Interlingüística* 14.681-694
- MATISOFF, JAMES, 1991. Areal and universal dimensions of grammaticalization in Lahu. *Approaches to Grammaticalization vol. II*, ed. by Elisabeth C. Traugott and Bernd B. Heine, 383-454. Amsterdam, Netherlands: John Benjamins Publishing Co.
- MAYORAL HERNANDEZ, ROBERTO. 2007. A variation study of verb types and subject position: verbs of light and sound emission. *Romance linguistics 2006: Selected papers from the 36th linguistic symposium on Romance Languages (LSRL), New Brunswick, March-April 2006, 219-233* ed. by José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez, and María José Cabrera, Amsterdam, Netherlands, John Benjamins Publishing Co.
- MEILLET, ANTOINE. 1911. L'évolution des formes grammaticales. *Scientia*. (Rivista di Scienza) 12(26).6. Reprinted in Meillet 1958: 130-48
- MELIA, J. (2003). *Modality*. Montreal: McGill-Queen's University Press.
- MEYERHOFF, MIRIAM. 2009. Animacy in Bislama? Using quantitative methods to evaluate transfer of a substrate feature. *Variation in indigenous minority languages*, ed. by James N. Stanford and Dennis R. Preston, 369-396. Amsterdam, Netherlands, John Benjamins Publishing Co.
- MILROY, LESLEY. 1987. *Language and social networks*. (2nd edition). Oxford: Basil
- MILROY, LESLEY and JAMES MILROY. 1992. Social network social class: Toward an integrated sociolinguistic model. *Language in Society* 21.1-26.
- MYHILL, JOHN. 1988a. The grammaticalization of auxiliaries: Spanish clitic climbing. *Berkeley Linguistics Society* 14.352-363.
- MYHILL, JOHN. 1988b. Variation in Spanish clitic climbing. *Georgetown University Round Table on Languages and Linguistics* 227-250.
- MYHILL, JOHN 1995. Change and continuity in the functions of the American English modals. *Linguistics* 33.157–211.
- MYHILL, JOHN. 1996. The development of the American English strong obligation system *American Speech*, 339-388.
- NARROG, HEIKO. 2005a. Modality, mood, and change of modal meanings: A new perspective. *Cognitive Linguistics* 16.677-731.

- NARROG, HEIKO. 2005b. On defining modality again. *Language Sciences* 27(2).165-192.
- NORDLINGER, RACHEL AND ELIZABETH CLOSS TRAUOGOTT. 1997. Scope and the development of epistemic modality: evidence from ought to. *English Language and Linguistics* 12.295-317.
- OCAMPO, FRANCISCO. 2006. Movement towards discourse is not grammaticalization: the evolution of *claro* from adjective to discourse particle in spoken Spanish. *Selected proceedings of the 9th Hispanic Linguistics Symposium*, ed. by In Nuria Sagarra and Almeida Jacqueline Toribio, 308-319. Somerville, MA: Cascadilla. Online: <http://www.lingref.com/cpp/hls/9/paper1388.pdf>
- OLBERTZ, HELLA. 1998. *Verbal periphrases in a functional grammar of Spanish*. Berlin: Mouton de Gruyter.
- PALMER, FRANK.R, 1979. *Modality and the English modals*. London, England: Longman.
- PALMER, FRANK. R. 2001. *Mood and modality*. Cambridge, England: Cambridge University Press.
- PAOLILLO, JOHN. 2002. *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: Center for the Study of Language and Information.
- PIERREHUMBERT, JANET. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137-157. Amsterdam, Netherlands: John Benjamins Publishing Co.
- PIETRANDREA, PAOLA. 2005. *Epistemic modality: functional properties and the Italian system*. Amsterdam, Netherlands: John Benjamins Publishing Co.
- PLANK, FRANS. 1984. The modals story retold. *Studies in Language* 83.305-364.
- POPLACK SHANA. 2001. Variability, frequency and productivity in the irrealis domain of French. *Frequency and the emergence of linguistic structure*, 405-428. ed. by Joan Bybee and Paul Hopper, 405–28.
- POPLACK, S., AND ELISABETE MALVAR. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus*, 19(1).121-169.
- POPLACK SHANA AND MARJORY MEECHAN, 1998. How languages fit together in codemixing. *International Journal of Bilingualism* 2(2).127-38.
- POPLACK, SHANA, and SALI TAGLIAMONTE. 2001. *African American English in the diaspora*. Malden, MA: Blackwell.

- RABADAN, ROSA. 2006. Modality and modal verbs in contrast: Mapping out a translationally relevant approach English-Spanish. *Languages in Contrast* 62.261-306.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) *Corpus diacrónico del español*. online: <http://www.rae.es>
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) *Corpus de referencia del español actual*. online: <http://www.rae.es>
- ROSSARI, CORINNE, CLAUDIA RICCI AND ADRIANA SPIRIDON. 2009. *Grammaticalization and pragmatics: Facts, approaches, theoretical issues*. Bingley, UK: Emerald Group Publishing.
- RUIZ DE MENDOZA IBÁÑEZ, FRANCISCO JOSÉ & LORENA PÉREZ HERNÁNDEZ. 2004. *High-level modal metonymies in English and Spanish*. *Jezikoslovlje 4; Special issue: How Universal are Conceptual Metonymies?* ed by Klaus-Uwe Panther and Linda L. Thornburg, , 103-120. Osijek, Croatia: University J. Strossmayer.
- SANKOFF, DAVID. 1988. Sociolinguistics and syntactic variation. *Linguistics: the Cambridge Survey*, 140-61. ed. by Frederick J. Newmeyer. Cambridge, England: Cambridge University Press.
- SCHNEIDER, EDGAR. 2002. Investigating Variation and Change in Written Documents. *The handbook of language variation and change*, ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 549-572. Malden, MA: Blackwell.
- SCHWENTER, SCOTT A. AND RENA TORRES CACOULOS. 2008. Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change* 20.1-39
- SKOTAREK, BEATA. 1996. The category of modality in Spanish. [Categoría de la modalidad en español] *Studia Romanica Posnaniensi* 21.91-114.
- SILVA-CORVALÁN, CARMEN. 1985. Modality and Semantic Change. *Historical Semantics. Historical Word Formation*, ed. by Jacek Fisiak, 547-572. Berlin: Mouton.
- SILVA-CORVALÁN, CARMEN. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford, England: Oxford University Press.
- SILVA-CORVALÁN, CARMEN. 1995. Interpretation of 'poder' and 'deber' *Modality in grammar and discourse*, ed. by Joan L. Bybee and Suzanne Fleischman, 67-106. Amsterdam, Netherlands: John Benjamins Publishing Co.
- SIRBU-DUMITRESCU, DOMNITA. 1988. Contribución al estudio de la semántica de los verbos modales en español con ejemplos del habla de Madrid. *Hispania* 71(1).139-148.

- SMITH, JENNIFER. 2000. Synchrony and diachrony in the evolution of English: evidence from Scotland. Unpublished PhD thesis, University of York.
- SMITH, JENNIFER, MERCEDES DURHAM & LIANE FORTUNE. 2007. "Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of a Scottish dialect. *Variation & Change*, 19(1).63-99.
- SMITH, NICHOLAS. 2003. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. *Modality in contemporary English*. ed. by Roberta Facchinetti, Manfred Krug and Frank Robert Palmer, 241-266. Berlin & New York: Mouton de Gruyter.
- SOLANO-ARAYA, JOSÉ MIGUEL. 1984. *Modality in Spanish: an account of mood*. Lawrence, KA: University of Kansas dissertation.
- SZMRECSANYI, BENEDIKT AND LARS HINRICHS. 2008. Probabilistic determinants of genitive variation in spoken and written English. *The dynamics of linguistic variation: Corpus evidence of English past and present*. ed. by Teritu Nevalainen, Irma Taavitsainen and Päivi Pahta, 291-310. Amsterdam; Netherlands: John Benjamins
- TAEYMANS, MARTINE. 2004. DARE and NEED in British and American Present-day English: 1960s-1990s. *New Perspectives on English Historical Linguistics. Vol. 1: Syntax and Morphology. An investigation into the emergence and development of the verb need* ed. by Christian Kay, Simon Horobin and Jeremy J. Smith, 215-228. Amsterdam: Benjamins.
- TAGLIAMONTE, SALI. 2002. Comparative Sociolinguistics. *The handbook of language variation and change* ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 729-763. Malden, MA: Blackwell.
- TAGLIAMONTE, SALI. 2004. Have to, gotta, must: Grammaticalisation, variation and specialization in English deontic modality. *Corpus approaches to grammaticalization in English*, ed. by Hans Lindquist and Christian Mair, 33-55. Amsterdam; Netherlands: John Benjamins.
- TAGLIAMONTE, SALI. 2006a. *Analysing sociolinguistic variation*. Cambridge, UK: Cambridge University Press.
- TAGLIAMONTE, SALI. 2006b. So cool, right?: Canadian English entering the 21<sup>st</sup> century. *The Canadian Journal of Linguistics*. 51(2).309-331.
- TAGLIAMONTE, SALI. and ALEXANDRA D'ARCY. 2007. The Modals of Obligation/Necessity in Canadian Perspective. *English World-Wide: A Journal of Varieties of English* 281.47-87.

- TAGLIAMONTE, SALI and JENNIFER SMITH, 2006. Layering, change and a twist of fate: Deontic modality in dialects of English. *Diachronica* 232.341-380
- TAGLIAMONTE, SALI, JENNIFER SMITH AND HELEN LAWRENCE. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17.75-112.
- TAGLIAMONTE, SALI, AND SHANA POPLACK. 1988. How black english past got to the present: Evidence from samana. *Language in Society* 17(4).513-533.
- TAGLIAMONTE, SALI, AND SHANA POPLACK. 1993. The zero-marked verb: Testing the Creole hypothesis. *Journal of Pidgin and Creole Languages* 8.171–206.
- TAGLIAMONTE, SALI AND SHANA POPLACK. 2000. The grammaticization of *going to* in (African American) English. *Language Variation and Change* 11:315-342
- TIPPETS, IAN. 2011. Differential object marking: Quantitative evidence for underlying hierarchical constraints across Spanish dialects. *Selected Proceedings of the 13<sup>th</sup> Hispanic Linguistics Symposium*, ed. by Ortiz-López, 107-117. Somerville, MA: Cascadilla Proceedings Project.
- TORRES CACOULLOS, RENA C. 1999. *Grammaticization, synchronic variation, and language contact: A study of Spanish progressive -ndo constructions*. Albuquerque, New Mexico: University of New Mexico dissertation.
- TORRES CACOULLOS, RENA. 2001. From lexical to grammatical to social meaning. *Language in Society* 30:443– 478.
- TORRES CACOULLOS, RENA. 2006. Relative Frequency in the Grammaticization of Collocations: Nominal to concessive *a pesar de*. *Hispanic Linguistics Symposium*, ed. by Timothy L. Face and Carol A. Klee, 8.37-49. Somerville, MA: Cascadilla Proceedings Project.
- TORRES-CACOULLOS, RENA. 2011. Grammaticalization and the Naturalization-in-Discourse Hypothesis. *The Handbook of Hispanic Sociolinguistics* ed. by in Manuel Díaz Campos, 148-167. Oxford, England: Blackwell.Publishing Ltd.
- TORRES CACOULLOS AND SCOTT A. SCHWENTER. 2004. Towards an operation notion of subjectification. Presented at the 31st Annual Meeting of the Berkeley Linguistics Society, California.
- TORRES CACOULLOS AND JAMES A. WALKER. 2009. On the persistence of grammar in discourse formulas: a variationist study of that. *Linguistics* 47.1-43.
- TRAUGOTT, ELIZABETH CLOSS. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 651.31-55.

- TRAUGOTT, ELIZABETH CLOSS. 1995. Subjectification in grammaticalisation. *Subjectivity and subjectification*, ed. by Dieter Stein and Susan Wright, 31–54. Cambridge, England: Cambridge University Press.
- TRAUGOTT, ELIZABETH AND RICHARD DASHER. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- TUCKER, THOMAS G. 1985. *Etymological dictionary of Latin*. Chicago, IL: Ares Publishers, Inc.
- URREA, EVELYN DURÁN, AND MICHAEL GRADOVILLE. 2006. *Variation in the Future Tense of New Mexican Spanish*. Albuquerque, NM: University of New Mexico.
- URRUTIA, HERNAN, AND MANUELA ALVAREZ. 1995. Haber and tener with the meaning of possession in medieval spanish. *RLA, Revista de lingüística teorica y aplicada* 33.209-21
- VAN DER AUWERA, J. AND V.A. PLUNGIAN. 1998. Modality's semantic map. *Linguistic Typology* 2.79-124.
- VAN DER GAAF, WILLEM. 1931. *Beon and habban* connected with an inflected infinitive. *English studies* 13.176-188.
- WÄRSNBY ANNA. 2004. Constraints on the interpretation of epistemic modality in English and Swedish. *English modality in perspective. Genre analysis and contrastive studies*, ed. by Roberta Facchinetti and Frank Palmer, 163-190. New York, NY: Peter Lang Publishing.
- WEINREICH, URIEL, WILLIAM LABOV, and MARVIN HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics: A Symposium* ed. by Windfred P. Lehmann and Yakov Malkiel, 95-188. Austin, TX: University of Texas Press.
- WESTNEY, PAUL. 1995. *Modals and periphrastics in English: An investigation into the semantic correspondence between certain English modal verbs and their periphrastic equivalents*. Tübingen: Max Niemeyer Verlag.
- WINFORD, DONALD. 1993. Variation in the use of perfect have in Trinidadian English: A problem of categorial and semantic mismatch. *Language Variation and Change* 5. 141–187.
- YLLERA FERNANDEZ ALICIA. 1980. *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Departamento de Filología Francesa, Universidad de Zaragoza.

## BIOGRAPHICAL SKETCH

Katherine Honea, raised in Fallon, NV, moved to California in 1996 to study. In 2001 she graduated with a Bachelor of Science in business administration with an emphasis in international marketing from California State University, Sacramento. She then worked for a couple years in her field as a sales representative before embarking on a year-long trip abroad in Chile. After spending a year in South America, and taking several prerequisite courses in California, she started a graduate program in Spanish at California State University, Sacramento. It was during this time she spent a semester and a summer studying in Mexico and Central America. After graduating in 2007 Katherine decided to pursue a doctoral degree in Spanish at the University of Florida, which she received in August 2012. While at the University of Florida, she traveled to Mexico to do field work. She also took a group of students to Santander, Spain for a six week long study abroad course as well as had the opportunity to develop and teach several upper-division content courses. She has presented her research at several national and international conferences.