FROM POSSESSION TO OBLIGATION: MODAL GRAMMATICALIZATION
AND VARIATION

A Dissertation in
Spanish
by
Joseph Bauman

© 2013 Joseph R. Bauman

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013
The dissertation of Joseph Bauman was reviewed and approved* by the following:

Rena Torres Cacoullos
Professor of Spanish and Linguistics
Dissertation Adviser
Chair of Committee

John Lipski
Edwin Erle Sparks Professor of Spanish and Linguistics

Philip Baldi
Professor emeritus of Linguistics and Classics

Richard Page
Associate Professor of German and Linguistics

Matthew J. Marr
Associate Professor of Spanish
Graduate Program Chair

*Signatures are on file in the Graduate School
ABSTRACT

Previous studies of grammaticalization and semantic change have found evidence of a crosslinguistic tendency for lexical verbs of possession to develop modal meanings (Bybee, Perkins and Pagliuca 1994; Heine and Kuteva 2002). The present research finds quantitative evidence of the mechanisms contributing to gradual meaning change in one such instance, that of the Modern Spanish deontic modal construction [tener que + Infinitive] ‘to have to’. In the process, the grammaticalizing form is considered as part of a variable context, in which it alternates with several other modal constructions.

The investigation outlined here is aimed at delineating and quantifying the processes at work via text-based, measurable evidence, and accomplishes two objectives. The first is to survey the grammaticalization of the [tener que + Infinitive] construction itself, assessing the factors involved its semantic development, such as semantic bleaching and abstraction, internal structural changes, increases in its token frequency, and its distribution among ever larger numbers of distinct lexical verb types. Second, the individual form is considered as part of a variable context, in which it alternates with several other modal constructions. Utilizing variationist methodology (Sankoff 1988) in the analysis of the construction’s diachronic advancement provides a theoretical and empirical perspective whose benefits have only begun to be explored by researchers of grammaticalization (Poplack 2011).

From a corpus of written texts and transcribed interviews representing Spanish from the 12th to the 20th centuries, a total of 5168 tokens were extracted. These were analyzed through an operationalization of various factors conditioning the occurrence of the tener que construction; the results find that grammatical person, aspect, and polarity are all influential in the selection of the innovative [tener que + Infinitive] over alternative forms (haber de, deber (de), haber que). In
the nine centuries considered here, the *tener que* construction develops from an effectively nonexistent particular usage of the lexical possession verb *tener* ‘to have’ to eventually become the predominant periphrastic modal expression used to express obligation. The systematic methodology used here to trace its development provides quantitative detail not found in previous accounts of the evolution of [*tener que* + Infinitive]. Furthermore, the results of this study confirm several predictions made by theoretical descriptions of grammaticalization, call into question certain claims about Spanish obligation expressions, and present a fresh perspective on the arrangement of conditioning factors that shape the modal domain of obligation.
# TABLE OF CONTENTS

List of Figures ......................................................................................................................... viii

List of Tables ........................................................................................................................... ix

Acknowledgements .................................................................................................................. xii

Chapter 1 Evidence of crosslinguistic tendencies in the grammaticalization of modality: possession and obligation ................................................................. 1

1.1. Characteristics and components of grammaticalization .................................................. 3
  1.1.1. Semantic change in grammaticalization: bleaching and abstraction ....................... 5
  1.1.2. Changes in frequency and distribution ..................................................................... 9
  1.1.3. Formal changes ........................................................................................................ 11
  1.1.4. Gradualness of change ............................................................................................ 12

1.2. Additional hypotheses ...................................................................................................... 13
  1.2.1. Modality .................................................................................................................. 14
  1.2.2. Polarity .................................................................................................................... 18
  1.2.3. Aspect ...................................................................................................................... 19

1.3. A crosslinguistic tendency: possession > obligation ...................................................... 20
  1.3.1. Previous studies ...................................................................................................... 21

1.4. The present study: design and methods .......................................................................... 23
  1.4.1. Chapter 2: Origins of the [tener que + Infinitive] construction: 12th-14th centuries .......................................................................................................................... 24
  1.4.2. Chapter 3: Semantic and structural changes in [tener que + Infinitive]: the 15th-18th centuries ................................................................................................................. 25
  1.4.3. Chapter 4: Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 19th century ......................................................................................... 26
  1.4.4. Chapter 5: Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 20th century ................................................................. 27

Chapter 2 Origins of the [tener que + Infinitive] construction: 12th-14th centuries ................. 29

2.1. The development of possession meaning ......................................................................... 30

2.2. ‘Hold, possess’-type uses of tener .................................................................................. 34
  2.2.1. ‘Hold’ vs. ‘possess’ .................................................................................................... 34
  2.2.2. Competition between tener and haber ...................................................................... 37
  2.2.3. The possessive/resultative construction with tener ............................................... 41
  2.2.4. ‘Stay’ ....................................................................................................................... 44
  2.2.5. ‘Extend’ ................................................................................................................... 44

2.3. tener for expressing beliefs and mental assessments ....................................................... 46
  2.3.1. ‘To believe’ .............................................................................................................. 48
  2.3.2. ‘To esteem’, ‘to regard as’ ....................................................................................... 49

2.4. Origins of the [tener que + Infinitive] construction .......................................................... 51
  2.4.1. Evidence of alternation with haber ......................................................................... 57
  2.4.2. Alternative constructions with tener .......................................................................... 58
  2.4.3. Semantic considerations .......................................................................................... 61

2.5. Conclusion ........................................................................................................................ 62
Chapter 3 Semantic and structural changes in [tener que + Infinitive]: the 15th-18th centuries .......................................................... 64

3.1. Changes in the frequency of [tener que + Infinitive] .................................................. 65
  3.1.1. Corpus and data extraction ............................................................................. 65
  3.1.2. Frequency of occurrence for [tener que + Infinitive] in the 15th-18th centuries .......................................................... 66
3.2. Semantic changes evident in [tener que + Infinitive] .............................................. 73
  3.2.1. Infinitive verbs frequently occurring within [tener que + Infinitive] ............. 75
  3.2.2. Bridging contexts ......................................................................................... 79
  3.2.3. tener que hacer .............................................................................................. 81
  3.2.4. tener que ver ................................................................................................. 85
  3.2.5. Semantic stages evident in the 15th-18th centuries ...................................... 88
  3.2.6. Infinitive types as evidence of semantic change .......................................... 91
3.3. Structural changes in [tener que + Infinitive] evident in the 15th-18th centuries ...... 94
3.4 Conclusion ........................................................................................................... 100

Chapter 4 Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 19th century .......................................................... 104

4.1. Alternative modal verb constructions for the expression of obligation ................... 105
  4.1.1. haber que + Infinitive ................................................................................. 107
  4.1.2. haber de + Infinitive .................................................................................. 110
  4.1.3. deber (de) + Infinitive ............................................................................... 111
4.2. Variation among the four alternative modal constructions: 12th-18th centuries ..... 113
4.3. Distribution of data in the 19th century ............................................................... 115
4.4. Productivity of [tener que + Infinitive] in the 19th century .................................. 121
4.5. Variationist analysis of the alternating modal variants .......................................... 124
4.6. Coding and factor groups .................................................................................. 125
  4.6.1. Grammatical person ................................................................................... 126
  4.6.2. Polarity ....................................................................................................... 128
  4.6.3. Tense/aspect .............................................................................................. 130
  4.6.4. Object presence and placement ................................................................. 132
  4.6.5. Source/genre .............................................................................................. 134
  4.6.6. Exclusions .................................................................................................. 135
4.7. Results ............................................................................................................. 136
  4.7.1. Individual analyses .................................................................................... 136
  4.7.2. [tener que + Infinitive] vs all three alternative variants ................................. 139
4.8. Conclusion ........................................................................................................... 141

Chapter 5 Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 20th century .......................................................... 144

5.1. 20th century corpus and data .............................................................................. 145
5.2. Distribution of data in the corpus ....................................................................... 147
  5.2.1. Written vs. spoken data ............................................................................. 151
  5.2.2. Iberian vs. Mexican data ............................................................................ 153
5.3. The occurrence of [tener que + Infinitive] in the 20th century ......................... 154
  5.3.1. Persistence of older meanings ................................................................. 157
5.3.2. The appearance of innovative modal meaning in \([tener que + Infinitive]\) ..... 159
5.3.3. Layering of meanings in \([tener que + Infinitive]\) ................................. 161
5.4. Multivariate analyses of the variation among the modal constructions: 20\textsuperscript{th}
century ................................................................. 163
5.4.1. Polarity ..................................................................................................................................................................................... 164
5.4.2. Aspect ...................................................................................................................................................................................... 165
5.4.3. Grammatical person ............................................................................................................................................................. 166
5.4.4. Medium and regional variety ............................................................................................................................................... 168
5.5. Results ..................................................................................................................................................................................... 168
5.5.1. Individual paired analyses .................................................................................................................................................. 169
5.5.2. Analysis of \([tener que + Infinitive]\) vs. all other variants ........................................................................................................ 171
5.6. Conclusion ................................................................................................................................................................................. 173

Chapter 6 Conclusions: Gradual development of \([tener que + Infinitive]\) and evidence of
grammaticalization processes ........................................................................................................... 177

6.1. Motivation for the study ............................................................................................................................................................ 177
6.2. Hypotheses regarding semantic changes ........................................................................................................................................ 178
6.3. Changes in frequency, distribution, and productivity: quantitative and
cronological evidence ........................................................................................................................................................................ 182
6.3.1. 12\textsuperscript{th}-14\textsuperscript{th} centuries .................................................................................................................................................. 182
6.3.2. 15\textsuperscript{th}-18\textsuperscript{th} centuries .............................................................................................................................................. 183
6.3.3. 19\textsuperscript{th} and 20\textsuperscript{th} centuries ......................................................................................................................................... 184
6.3.4. Multivariate analyses ............................................................................................................................................................. 186
6.4. Implications for theoretical descriptions of grammaticalization processes .... 187
6.5. The path from possession to obligation ........................................................................................................................................ 189
6.6. Future investigation .................................................................................................................................................................... 192

References ............................................................................................................................................................................. 195

Appendix Corpus of texts and transcribed interviews .......................................................... 204
12\textsuperscript{th}-14\textsuperscript{th} centuries (Chapter 2) ........................................................................................................................................ 204
15\textsuperscript{th}-18\textsuperscript{th} centuries (Chapter 3) ........................................................................................................................................ 205
19\textsuperscript{th} century (Chapter 4) ............................................................................................................................................... 205
20\textsuperscript{th} century (Chapter 5) .............................................................................................................................................. 207
LIST OF FIGURES

Figure 1-1. Proposed stages of semantic and structural evolution of English *have to*..............22

Figure 2-1. Developmental stages in grammaticalization of English *have to* + INFINITIVE ....52

Figure 3-1. Frequency of occurrence of *[tener que + Infinitive]* in the selected corpus of
texts from the 12th to the 18th century (number of tokens per million words)...............69

Figure 3-2. Frequency of occurrence of *[tener que + Infinitive]* in the supplemental
corpus of texts from the 15th to the 18th century (number of tokens per million
words) ..........................................................................................................................71

Figure 3-3. Overall frequency of occurrence of *tener* ‘to have’ in Davies’ (2002-) corpus
from the 13th to the 18th century (number of tokens per million words) .......................72

Figure 3-4. Developmental stages in the evolution of the English *to have to* construction
(drawn from Brinton 1991 and Fisher 1994) ..........................................................................88

Figure 3-5. Relative frequencies of three different object types from the 15th-18th
centuries in *[tener que + Infinitive]* ..................................................................................97

Figure 4-1. Percentages of data accounted for by four modal constructions expressing
obligation from the 12th to the 18th century ........................................................................114

Figure 4-2. Relative frequencies of modal constructions expressing obligation from the
17th to the 19th century ..........................................................................................................118

Figure 4-3. Percentage of total data accounted for by the four most frequent lexical
infinitives in each century from the 15th to the 19th ..............................................................124

Figure 5-1. Relative frequencies of modal constructions from the 17th to the 20th ..........149

Figure 5-2. Relative frequencies of modal constructions from the 18th to the 20th century,
distinguishing written data from spoken data ....................................................................152

Figure 6-1. Relative frequency of *[tener que + Infinitive]* with respect to three alternative
obligation constructions ([*haber de / deber (de) / haber que* + Infinitive]) ....................186

Figure 6-2. Token counts for *[tener que + Infinitive]* and *[haber que + Infinitive]* from
the 18th to the 20th century ..................................................................................................191
LIST OF TABLES

Table 2-1. Corpus of 12th-14th century texts. .................................................................33
Table 2-2. Metaphorical uses of tener by century. ...............................................................51
Table 2-3. Token counts for each use of tener by century....................................................53
Table 3-1. Corpus of 15th-18th century texts and [tener que + Infinitive] tokens extracted
from each.........................................................................................................................68
Table 3-2. Supplemental corpus and data from Davies’ (2002-) Corpus del Español ..........70
Table 3-3. Distribution of the most common infinitive types for the 15th-18th centuries.......75
Table 3-4. Distribution of common infinitive types for the 15th-18th centuries in Davies’
(2002-) corpus ................................................................................................................77
Table 3-5. Type/token frequencies for infinitives in [tener que + Infinitive] and
proportion of data composed of the four most frequent infinitives (dar, decir, hacer, ver) in the intermediate time period .................................................................92
Table 3-6. Percentage of tokens of [tener que + Infinitive] with postposed objects from
the 15th to the 18th century...............................................................................................95
Table 3-7. Percentage of tokens of [tener que + Infinitive] with intervening objects from
the 15th to the 18th century...............................................................................................98
Table 4-1. Token counts for constructions encoding obligation for the 12th-18th centuries.....113
Table 4-2. Corpus of 19th century texts...............................................................................116
Table 4-3. Token frequencies of periphrastic expressions of obligation in the 18th and 19th
centuries (in number of tokens per million words) .........................................................117
Table 4-4. Type/token ratios for lexical infinitive types appearing in [tener que +
Infinitive] from the 15th to the 19th century.................................................................121
Table 4-5. 19th century tokens of the modal variants by grammatical person and number
(percentage of total tokens comprising each variant in parentheses)...............................128
Table 4-6. 19th century negation rates for four variant modal constructions ......................130
Table 4-7. Distribution of modal construction tokens based on tense/aspect ......................132
Table 4-8. Results of multivariate analyses of factors contributing to choice of [tener que +
Infinitive] as opposed to each of the alternative variants in the 19th century. Factor
weights indicate the contribution of factors selected as significant; brackets indicate
that a factor was not selected as significant in a given analysis........................................137
Table 4-9. Results of multivariate analysis of factors contributing to the selection of [tener que + Infinitive] over [deber (de)/haber de/haber que + Infinitive] ..................140

Table 5-1. Corpus of 20th century texts and interviews (authors for written works in parentheses).......................................................................................................................147

Table 5-2. Obligation tokens extracted from the 20th century corpus..............................................148

Table 5-3. Frequencies (occurrences per million words) of the four alternative modal constructions in the 18th-20th centuries........................................................................................................150

Table 5-4. Frequencies (per million words) of the four modal variants in the 20th century: written language vs. spoken language.....................................................................................................151

Table 5-5. Frequencies (per million words) of the modal variants in the 20th century, arranged by variety.........................................................................................................................153

Table 5-6. Type/token ratios for [tener que + Infinitive]: 15th century to 20th century (ratio of number of distinct lexical types appearing in the construction to total number of tokens) .........................................................................................................................155

Table 5-7. Relative clause (possession) occurrences of [tener que + Infinitive] vs. obligation tokens in the 20th century corpus.............................................................................................................158

Table 5-8. Distribution of 20th century tokens by polarity (percentage of tokens comprising each variant by polarity in parentheses).......................................................................................164

Table 5-9. Distribution of 20th century data by aspect........................................................................165

Table 5-10. Distribution of 20th century tokens by grammatical person/number .........................167

Table 5-11. Results of multivariate analyses pairing [tener que + Infinitive] respectively with [deber (de) + Infinitive] and [haber que + Infinitive] in the 20th century.........................170

Table 5-12. Results of multivariate analysis of factors contributing to the selection of [tener que + Infinitive] over [deber (de)/haber de/haber que + Infinitive]: 20th century..............................................172
DEDICATION

To Max and Eddie: my memory of the past, my strength in the present,
and my hope for the future.
ACKNOWLEDGEMENTS

A number of individuals have contributed to making this work what it is. Trying to list them individually while limiting my account of their assistance to a single page seems to undervalue what they have done, as their support has been indispensable. Rena Torres Cacoullos, my adviser, has been a constant source of good counsel and strategic guidance for several years. Her suggestions have not only influenced the direction of this research, but also ensured its fruitful outcome; she has always steered me in the right direction. I am also grateful to my committee: since I began my graduate program, John Lipski has provided an example of a true linguist, scholar, and intellectual adventurer; Richard Page has added a valuable perspective on language change outside of Romance; Phil Baldi, erudite in all things Indo-European, became a much-needed mentor years ago when I was enthusiastic but “rudderless”.

Long before this investigation coalesced, my parents Don and Kathy equipped me with my desire to explore, improve, and ask questions. Without that will to grow and succeed, I would not be where I am now, nor would I be able to look forward to where I may go. They kindled the spark of my relentless pursuit of knowledge, which ultimately led me to turn an interest into an occupation. My wife Jamie, a constant and inspiring presence in my life during a time of change and uncertainty, has given me what is perhaps the most important resource for completing an endeavor such as this: motivation. She knew what I needed to keep me going even when I didn’t know it myself. Without her encouragement, which was sometimes conveyed wordlessly in a hug and sometimes scrawled in marker on the bathroom mirror, this dissertation may not exist at all.
Chapter 1

Evidence of crosslinguistic tendencies in the grammaticalization of modality: possession and obligation

For much more than a century, researchers of linguistic change have sought to identify and explain diachronic processes that recur in numerous languages. Considering evidence from various language families, scholars have identified tendencies that suggest universal mechanisms of change. In the past few decades, many of these trends have been classified as examples of grammaticalization, a set of processes by which lexical items in particular constructions acquire grammatical meaning (Bybee 2010: 106) The research program outlined here targets one such change, the transformation of a verb that encodes possession into a marker of obligation.

In Modern Spanish, the construction [tener que + Infinitive] is used to express obligation, but its origin lies in a particular use of the lexical verb tener ‘to have’, a form that continues to encode possession in the contemporary language. Compare the examples in (1-2).

(1) Lo tengo yo. Es sensacional ese libro.  
   It have.PRS.1SG I be.PRS.3SG sensational that book
   ‘I have it. That book is sensational.’ [20th C.; MadCult: 449]

(2) lo tengo que hacer, porque es necesario  
   it have-to.PRS.1SG do.INF because be.PRS.3SG necessary
   ‘I have to do it, because it is necessary.’ [20th C.; MexCult: 26]

The semantic development from possession to obligation is attested in several language families but has been discussed with only limited detail. The English manifestation of the change (have >
have to) has been addressed the most thoroughly (Brinton 1991; Fischer 1994; Krug 2000; Dollinger 2006), but few of these studies approach the process using quantitative evidence. Descriptions of the Spanish example with tener (e.g. Yllera 1980) also tend not to be systematic treatments of the change.

The data and analyses presented in the following chapters are aimed at investigating the grammaticalization of the tener que construction in a more scientific, detailed way than have existing accounts of the change. A corpus of written texts and transcribed interviews is compiled here from which are extracted all occurrences of the target construction over a period of nine centuries (the 12th to the 20th). Via the analysis of these data, the research outlined herein will seek to accomplish two goals. The first is to investigate [tener que + Infinitive] as a case study for the acquisition of modal meaning by a possession verb and provide a thorough chronological description of its evolution. The second goal is to place that specific example in a larger context, using it to attain text-based, measurable evidence of the linguistic machinery that is theoretically assumed to drive the general phenomenon of grammaticalization.

As an introduction to the matter, this chapter is arranged as follows: first, an explanation of grammaticalization itself is presented, pointing out the testable hypotheses offered by the extant theoretical literature as well as the shortcomings of the same. Second, the specific path of grammaticalization from possession to obligation traced by the tener que construction is discussed, as well as the aspects of modality and the set of additional Spanish modal constructions that are involved in its development. The information presented in these two sections is then used to generate hypotheses to be tested in the research project. Finally, the study itself is outlined, including an explanation of the data and methods.
1.1. Characteristics and components of grammaticalization

Central to the investigation outlined here is the phenomenon of linguistic change known as grammaticalization\(^1\). The earliest mention of the phenomenon in its modern sense is commonly attributed to Meillet, who introduced the term *grammaticalisation*, referring to the general tendency for some independent lexical items to take on grammatical functions (le passage de mots autonomes au rôle d’agents grammaticaux; 1912: 133). He defined this notion as a weakening of the pronunciation, the concrete sense, and the expressive value of words. Advanced semantic “weakening” will, according to Meillet, leave a form without an independent meaning of its own, causing it to rely on another word or words to indicate its grammatical role.

More recent authors provide similar definitions of the phenomenon. Heine and Reh call it “an evolution whereby linguistic units lose in semantic complexity, pragmatic significance, syntactic freedom, and phonetic substance […].” (1984:15). As an example of this bundle of changes, the English verb *go* became fused to the preposition *to* in the *[be going to + Infinitive]* construction, losing its lexical sense of physical motion and instead encoding future temporal reference. In the process, what were two elements became one, allowing for phonetic reduction of the form concomitant with an increase in its overall frequency, evident in the form’s common rendering as *gonna* /gʌnə/.

Even prior to Meillet’s explicit adoption of the term “grammaticalization”, authors as early as the 19\(^{th}\) century had alluded to general tendencies in linguistic change. In most cases, these authors noted the development of closed-class (grammatical) categories from open-class (lexical) items as a general pattern (see especially Heine, Claudi and Hünnemeyer 1991: 5-8 and the citations therein). Near the end of the century, the neogrammarians Georg von der Gabelentz synthesized the existing evidence and framed the process as a sort of evolutionary paradigm. He

\(^1\) The term “grammaticalization” is at times use interchangeably with “grammaticization” (Bybee et al. 1994)
suggested that linguistic forms were like employees of the state who were first hired, then
promoted, placed on reduced pay, and ultimately retired as new applicants lined up to replace
them (Hopper and Traugott 1993: 19-20).

Following the turn of the 20th century and Meillet’s adoption of a specific term for the
tendency, grammaticalization began to be viewed as a tool to be utilized in diachronic linguistics
for reconstructing a history for a given language or group of languages. This perspective was
presumably shaped by the desire to identify typological universals, especially with respect to
word order (Greenberg 1966, for example). Even so, it was apparently still not considered a large-
scale instrument on par with a teleological notion like drift (Sapir 1921, Vennemann 1975).

Eventually, grammaticalization began to be used as an explanatory device for particular
elements of diachronic change and even to account for certain synchronic components of
linguistic organization, given its link to word order and syntactic structure (e.g. Givón 1975 and
1978; Li and Thompson 1974). This new perspective is perhaps best summed up by Givón’s
assertion that “today’s morphology is yesterday’s syntax” (1971: 413). From there,
grammaticalization gained traction as a topic of inquiry in its own right, eventually being
assessed as the central theme in works like those of Lehmann (1982) and Traugott (1982).

The mantle of grammaticalization has been assumed in the past few decades by authors
like Bernd Heine, Elizabeth Closs Traugott, Paul Hopper, and Joan Bybee, who have individually
and collectively created numerous works aimed specifically at examining grammaticalization,
including eponymous volumes on the subject (e.g. Heine et al. 1991; Hopper and Traugott 1993).
Even more recently, grammaticalization has been coupled with cognitive (Langacker 1987) and
usage-based approaches to grammar (Bybee 2006) to form an all-encompassing theoretical
perspective, marrying both the synchronic state of a language and the diachronic developments
that have brought it about (see especially Bybee 2010). As such, grammaticalization theory
makes certain predictions about how changes unfold, which themselves can be utilized as testable
hypotheses. The following sections outline those predictions that are most pertinent to the present examination of the [tener que + Infinitive] construction.

1.1.1. Semantic change in grammaticalization: bleaching and abstraction

Heine et al. (1991: 14) credit Traugott with making semantic change a principal component, and perhaps the primary focus, of much of the recent research on grammaticalization. Indeed, a crucial mechanism in all accounts of grammaticalization is the notion of semantic loss, sometimes called “bleaching”. Semantic bleaching is sometimes even considered to be effectively synonymous with grammaticalization itself (Heine et al. 1991: 3). In their treatise on grammaticalization, Hopper and Traugott note that Gabelentz (1891) was among the first linguists to allude to “bleaching” in connection with linguistic change. He stated that linguistic forms “fade, or grow pale” and that “their colors “bleach”, needing to be covered over with a fresh coat of paint” (Hopper and Traugott 1993: 20).

Certainly tener in the tener que construction exhibits some loss of semantic content when examples of the construction like that in (3), from the 17th century, are viewed alongside later examples, such as the 19th century token in (4).

(3) Déjame, Tristán, que tengo que hacer.
leave.IMP.3SG-ACC.1SG Tristán COMP have.PRS.1SG that-which do.INF
‘Leave me Tristán, as I have things to do.’ [17th c.; Davies 2002-]

(4) pero tengo que hacerte una confesión muy importante.
But have-to.PRS.1SG make.INF-DAT.2SG a confession very important
‘But I have to make a very important confession to you.’ [19th c.; Davies 2002-]
In (3), tener refers to an object (chores or a responsibility) that is possessed, albeit in an abstract sense. But in the later token in (4), the reading is one of obligation rather than possession; the confesión ‘confession’ is not something possessed. Instead, it refers to action the speaker has not yet performed, but is obligated to do so.

The change illustrated by (3-4) can undoubtedly be considered one of loss. The original sense of possession inherent to tener ‘to have’ is absent in (4). However, it is important to note that the modal meaning (obligation) has been added where it was not present before. Such a change could be viewed as one of abstraction, rather than loss. This is an alternative explanation of the semantic change endemic to grammaticalization offered by Heine et al. (1993: 41), who suggest that grammatical meaning can be considered more abstract (less concrete) than lexical meaning. Comparing possession to obligation, it could be argued that the latter is relatively more concrete than the latter, and rather than necessarily losing semantic content, the possession meaning has been abstracted to obligation.

As part of the research presented here, it will be necessary to determine whether the semantic changes evident in the grammaticalization of the [tener que + Infinitive] construction are best characterized as bleaching, abstraction, or some other option perhaps incorporating elements of both. Furthermore, the question to be answered will focus on what, if any, quantitative evidence there is of the semantic changes undergone by the construction as it takes on modal meaning.

1.1.1.1. Semantic persistence

Pertinent to semantic change in grammaticalization is the claim by Hopper that as a form grammaticalizes, while some older elements of its meaning may be lost, “some traces of its original lexical meaning tend to adhere to it, and details of its lexical history may be reflected in
Given that tener ‘to have’ is a possession verb, and indeed continues to be the predominant form used to encode possession in Modern Spanish, it is possible that bits of its lexical meaning will persist as the [tener que + Infinitive] construction grammaticalizes. To search for evidence of this persistence, it will be necessary to begin by investigating the occurrence of the construction with possessed objects and presumably transitive infinitive verbs, as in (5).

(5) et non teneremos ninguna cosa que comer

‘and we do not have anything to eat’ [13th c.; Calila: 326]

In this case ninguna cosa ‘not a thing’, is the possessed object of tener ‘to have’, while at the same time acting as the patient of the transitive infinitive comer ‘to eat’, which represents an action to be performed on the object. As [tener que + Infinitive] grammaticalizes, the inclusion of objects and transitive lexical infinitives may play a role in determining its distribution, as suggested by Hopper. The research project outlined in Section 1.4 includes a focus on these elements in its design.

1.1.1.2. Layering

Another assertion by Hopper holds that within a functional domain, new layers of meaning continuously emerge and “as this happens, the older layers are not necessarily discarded, but may remain to coexist and interact with the newer layers” (1988:22). As an example, a periphrastic future construction in Latin (cantāre habet ‘he has to sing > he will sing’) coexisted temporarily with a synthetic future of the type cantabit ‘he will sing’ before the former eventually
replaced the latter and became fused as a synthetic form itself: *cantará* ‘he will sing’ in Modern Spanish (Fleischman 1982: 50-66; Fruyt 2009: 800-810). Thus, rather than filling a gap in a particular functional domain, the emerging form is increasingly applied to a function that is already presumed to be part of the domain of the earlier form. The two may not be semantically identical at first (perhaps only overlapping in a particular context), but the innovative form gradually “extends its range of meanings to include those of the construction which it replaces” (Hopper and Traugott 1993: 125).

The notion of layering is consonant with what scholars of linguistic variation have long called inherent variability in a domain for which multiple forms can be argued to express the same meaning (Labov 1969; Sankoff 1988a, inter alia). Furthermore, it is no new proposal that a comprehensive theory of language change should be based on empirical foundations (Weinreich, Labov and Herzog 1968). The interface between the investigation of grammaticalization and variation theory is put succinctly by Poplack, who argues that “extending the focus to the entire variable context confirms that forms do not grammaticalize in isolation, but are sensitive to the variants with which they alternate” (2011: 222-223). Thus, if we are to pursue a detailed account of how a given form grammaticalizes, it is prudent to examine the variable context in which it grammaticalizes, i.e. the forms with which it varies (see Sections 1.4.3 and 1.4.4).

In the history of Spanish, an opportunity to consider a grammaticalizing form in its variable context presents itself in the domain of obligation. Two periphrastic modal constructions predated [*tener que* + Infinitive] in the domain, each of which incorporated a modal verb and a lexical infinitive, as in (6-7)².

---

² The present study is limited to the periphrastic constructions featuring a modal verb + lexical infinitive combination that have survived as expressions of obligation in the modern language. There were other forms (e.g. *ser menester* ‘to be necessary), but many of these fell into disuse several centuries ago (see Yllera 1980 for a comprehensive list).
[**haber de** + Infinitive]

(6) *et por cuyo consejo fazían lo-que avían de fazer*

and by POSS advice do.IPFV.3PL that-which have.IPFV.3PL of do.INF

‘and by his advice they did what they had to do’ [13th c.; Calila: 225]

[**deber (de) + Infinitive**]³

(7) *Et dizên que el omne entendido deve contar a su padre et a su madre por amigos*

‘and they say that the wise man should count his father and mother as friends’

[13th c., Calila: 274]

When compared with examples of the *tener que* construction like those in (2) and (4) above, it is clear that these forms can be considered distinct layers of the same general meaning, or variants of a linguistic variable. The research presented here is aimed at assessing the extent to which these older forms overlap chronologically with the innovative *tener que* construction and characterizing the variation between them in the functional domain of expressing obligation.

### 1.1.2. Changes in frequency and distribution

Almost as indispensable in defining grammaticalization as the semantic changes described above is the frequency with which a given form occurs. Researchers of grammaticalization tend to be in agreement that the frequency of a grammaticalizing form is an essential component of the process. However, the role played by frequency is evidently not clear-cut. On one hand, Heine et al. suggest that the linguistic forms serving as the source of grammaticalization tend to be “of frequent and general use” (1991: 38). This contrasts with what

---
³ The *de* element is placed in parentheses here to indicate that is presence in the construction is variable. The reason for doing this is explained in Section 4.1.3.
Hopper and Traugott argue: “given that frequency demonstrates a kind of generalization in use patterns”, “the more frequently a form occurs in texts the more grammatical it is assumed to be.” (1993:103).

The circularity of these arguments is immediately apparent, but that does not mean both assertions cannot be true. Bybee reconciles the two views by pointing out the link between frequency and each component process included in grammaticalization:

…because all the processes depend in one way or another upon repetition, increases in frequency trigger their operation, while at the same time the output of the processes (semantically more generalized meanings or a wider applicability due to inferences) leads to further frequency increases. (2010: 112-113).

Returning to the example of English going to mentioned in Section 1.1, it is possible to see increased frequency of the construction’s occurrence as both a cause and effect. As going loses its lexical sense of physical motion, the construction can be generalized to appear in contexts where motion cannot be a possible interpretation: compare the ambiguity in I am going to work with the clear future reference in I am going to stay here. Generalizing to new contexts leads to an increase in frequency, which in turn increases the probability that going to will be reduced to gonna (see Bybee 2010: 37-41 for an overview of the reducing effect of frequency).

Bybee’s statement about the role of frequency is important to analyzing the evolution of [tener que + Infinitive] for two reasons. First, frequency is a readily quantifiable variable, though there are different measures, namely token and type frequency, or frequency relative to another variant. Assuming that it can be interpreted as evidence of grammaticalization, measuring the frequency with which [tener que + Infinitive] occurs will be a key indicator of the construction’s status. Bybee also mentions “generalized meanings” and “wider applicability”, which in the case of the tener que construction will impact the contexts in which the construction appears. As
mentioned above in reference to semantic persistence, it will be necessary to track the range of lexical verbs appearing in \( [tener \ que \ + \ \text{Infinitive}] \) to determine what changes are taking place in its distribution in the grammar.

### 1.1.3. Formal changes

It is often noted in writings on grammaticalization that in advanced stages, grammaticalizing forms undergo a loss of compositionality or analyzability (Bybee 2010:50-51). That is, what was once analyzable as having a complex internal structure is more likely to be analyzed as a simplex form. This is particularly evident in the reduction of going to, as gonna is not transparently divisible into multiple morphemes. In many cases, increased internal bondedness between what were multiple elements in a complex form can actually arrive at a synthetic form (Hopper and Traugott 1993: 8-9) The transition from cantāre habet to cantará, where the former is periphrastic and the latter bears only the affixal –á as a remnant of what began as a form of habēre, illustrates this change.

However, increased bondedness within a construction need not only be identifiable in the transformation of an analytic form to a synthetic form. Bybee argues that even low degrees of grammaticalization can exhibit evidence of loss analyzability (2010:51). As \( [tener \ que \ + \ \text{Infinitive}] \) remains periphrastic in Modern Spanish, a potential indicator of this is the presence and/or placement of a direct object in the construction\(^4\). Examples like (8), where an overt object (dineros ‘money’) intervenes between tener and que would suggest that those two forms are not analyzed as a single unit. In contrast, (9) bears no overt object despite including the transitive lexical verb pagar ‘to pay’, which would presumably entail an implicit object (money).

\(^4\) For additional discussion of the placement of object pronouns as an indication of the status (e.g. analytic vs. synthetic) of occurrences of the grammaticalizing Spanish future construction mentioned in the previous paragraph, see Company Company (2006: §4.6.4).
(8) *Yo no tengo dineros que os dar por la llave* [16th c.; *Lazarillo*: 26]

‘I do not **have money to give** you for the key’

(9) *compré una recua y unas casas, y e tenido que pagar de ay más que quería*

‘I bought a mule train and some houses, and I have **had to pay** therefore more than I wanted to’ [16th c.; *DLNE*: 220]

The presence of intervening material between *tener* and *que* in the construction can be taken as an indication of relatively less advanced grammaticalization, assuming that a greater degree of grammaticalization will correlate with those two elements being adjacent.

1.1.4. Gradualness of change

A final hypothetical point about grammaticalization that can be tested is the assertion that it is necessarily gradual. This is one area in which grammaticalization as a theoretical perspective differs fundamentally from a traditional generative approach to language change. Rather than assuming the parameters that make up a given language’s grammar are abruptly reset or that a new form directly replaces an older one, grammaticalization theory takes as a given that earlier meanings coexist with innovative meanings and older forms coexist with newer ones (see Hopper and Traugott 1993:35-37 for a more detailed discussion of this distinction). The examples in (8-9) illustrate this coexistence of meanings in the history of the *tener que* construction. While (8) indicates possession and (9) suggests obligation, both are drawn from 16th century texts.

The methodological approach adopted here is particularly well-suited to detect this sort of evidence of gradual change. Using evidence based on written texts dating from the 12th to the 20th century, it is possible to determine the extent to which older meanings persist and coexist with
newer ones as [tener que + Infinitive] grammaticalizes. Furthermore, the chronological range will capture the duration of that coincidence. And as mentioned in Section 1.1.1.2, a comparison of the tener que construction’s distribution with those of functionally equivalent, but older, constructions is also included in the investigation.

1.2. Additional hypotheses

In addition to general observations about how the processes of grammaticalization unfold, a number of authors have made more specific observations that are pertinent to the topic addressed here. As with the broader issues discussed in Section 1.1, these observations can be adopted as hypotheses regarding the grammaticalization of [tener que + Infinitive]. They pertain to the grammaticalizing form in its variable context, broadening the scope of investigation to include a system of forms instead of a single form itself. That is, rather than making predictions about the tener que construction’s development per se, these observations are better suited to accounting for the variation between [tener que + Infinitive] and the alternative modal constructions that are assumed to share the general domain of obligation in Modern Spanish.

These constructions include the two noted above in examples (6) and (7) with haber de and deber (de) as well as a third, [haber que + Infinitive], which developed comparatively recently in the history of the language, as did [tener que + Infinitive]:

(10) Ya ve usté que hay que hacer el quihacer a la hora que se necesita
    ‘so you see that one must carry out the chore when it is necessary’

[20th c.; MexPop: 171]
The \textit{tener que} + Infinitive] construction has coexisted with the three other obligation constructions for several hundred years and has overlapped with them in the expression of obligation especially since the 19th century. A thorough treatment of the variation between the four forms is conducted in Chapters 4 and 5, where are tested the hypotheses drawn from observations made in Sections 1.2.1-1.2.3 below.

1.2.1. Modality

A focal point of the study presented in the following chapters is the modal domain of obligation. Thus, the analyses included therein and the conclusions drawn from them depend on a few predetermined distinctions, given the complex and nebulous nature of modality. It is often difficult, for example, to categorize a given token of a modal form as unequivocally expressing either obligation (deontic modality) or probability, an example of epistemic modality.\(^5\) Consider the occurrences of \textit{tener que} + Infinitive] in (11-12):

(11) \textit{Figúrese usted que recibe el susto, y que además le quedan algunos huesos quebrantados, sin nada de heridas graves, se entiende... pues en tal caso, o se acobarda y huye de Orbajosa, o se tiene que meter en la cama por quince días.}

‘Imagine that he gets frightened, and in addition he has a few bones broken, without any serious wounds, of course…Well in that case, either his courage will fail him and he will leave Orbajosa, or he will \textbf{have to keep} to his bed for fifteen days.’ [19th c.; \textit{Doña Perfecta}: 200]
In this instance, it is easier to argue that (11) presents an example of deontic (or root) modality, whereas (12) depicts an epistemic usage. The former, wherein the assertion tiene que meter en la cama ‘he has to stay in bed’ is based on objective evidence might be more easily classified as expressing obligation. That is, in order to recover from his injuries, the subject will require a certain period of convalescence. In (12), the proposition tenía que ser él ‘it must be him’ depends on a supposition: the individual who opened the door could not have been certain of who was on the other side of it, but assumed that she did.

In this instance, it is relatively easy to categorize the two tokens. Even so, it could be argued that rather than each belonging to one discrete class or another, the tokens in (11-12) occupy two positions on a continuum. Both exhibit propositions that depend on external knowledge (familiarity with how long it might take for someone to heal vs. expecting a certain visitor at a given time), and both could be argued to be conjectures (neither has necessarily been proven to be true). The difference is that one is relatively more deontic (11), while the other is relatively more epistemic (12).

In her description of the semantic content of modal auxiliaries in British English, Coates acknowledges that at times it is useful to conceive of modal meanings as falling into distinct categories while in other cases they appear to be more continuous. She writes: “neither models which assume discrete categories nor those which assume indeterminacy are wholly satisfactory for an analysis of modal meaning” (1983: 10). For this reason, Coates marries the two perspectives and utilizes a “fuzzy set” approach to account for various modal meanings. She argues that a given meaning (e.g. obligation) can be assumed to belong to a category, a “core”
meaning, which extends outward into a “skirt” including less central meanings and then into an even more tangentially related “periphery” (Coates 1983: 12-13).

It is useful to approach the topic of \textit{tener que + Infinitive} acquiring modal meaning from this viewpoint. As has been mentioned, grammaticalization is assumed to entail gradual change and it must be acknowledged that the boundaries between the various meanings of the \textit{tener que} construction are not always immediately apparent. But for the sake of synthesizing results and making generalizations, it is necessary to classify individual occurrences of the construction as belonging to one category or another. For instance, despite their overlap, the token of the \textit{tener que} construction in (11) would be counted as encoding obligation, whereas (12) would be counted as expressing probability.

1.2.1.1. Subjectivity and subjectification

The distinction between (11) and (12) could be described as a difference in subjectivity, or the speaker’s “perspective…or point of view in discourse” (Finegan 1995: 1). Thus, it concerns the linguistic manifestation of the speaker’s attitudes toward what is being said. As Stubbs puts it, “whenever speakers (or writers) say anything, they encode their point of view towards it” (1986:1). By implication, subjectivity is commonly associated with the development of epistemic modal meaning.

Langacker (1990) asserts that the distinction between root (deontic) modality and epistemic modality depends on the source of the imposed obligation on an agent to perform an action. For a deontic usage, obligation is viewed as being imposed by the speaker, another person, a law, a social norm, or cultural expectations. This applies to (11), where the speaker’s statement that ‘he will have to keep to his bed for fifteen days’ is based on the inevitability of taking time to
recover from injuries. In contrast, the example in (12) includes a speaker’s assumption based on her knowledge of reality itself, rather than coming from a specific entity: ‘it had to be him’.

Important for the present research is Langacker’s argument that “extension from a root to an epistemic meaning constitutes a type of subjectification” (1990: 28). Traugott agrees, stating that a cross-linguistic tendency of semantic change is for meanings to become increasingly based in the speaker’s subjective belief state or attitude toward a proposition (1989: 35). Furthermore, Traugott and Dasher posit that the most subjective expressions will include explicit markers of the speaker’s (or writer’s) “attitude to what is said, including epistemic attitude to the proposition (2002: 23).

Two of the constructions with which [tener que + Infinitive] alternates in the modal domain of obligation (those with haber de and deber (de)) are more commonly associated with expressing probability, a sense of epistemic modality, than is tener que (Butt and Benjamin 2004: 328-330). Given that these two constructions predate the tener que construction as modal forms and have moved further into the domain of probability, it can be hypothetically presumed that their occurrence as epistemic constructions will be correlated with greater subjectivity than that of [tener que + Infinitive].

However, in the domain of obligation, the tener que construction is the innovative form and is actively grammaticalizing, acquiring modal meaning during the time period examined here. With this in mind, occurrences of the tener que construction are expected to be associated with subjectivity, in this domain. Traugott states that the expression of a speaker’s belief or attitude toward what is said will be “grammatically identifiable” as a form grammaticalizes (1995: 32), and that subjectivity is operationalized here as part of the analyses of conditioning factors contributing to the choice of [tener que + Infinitive] over the modal variants with which it alternates in the domain of obligation (see Section 4.6.1).
1.2.2. Polarity

A variable that is potentially linked with subjectivity is the polarity of a clause. In his discussion of the pragmatic and functional role of negation, Givón argues that “the ontology of negation in language is firmly grounded in the perceptual principle of figure/ground” where negative contexts are more marked relative to their corresponding affirmatives (1978: 109). That is to say, a speaker’s reporting a negative event is contrary to a hearer’s default presupposition of the affirmative action. According to Givón, negation therefore makes a given event into a ground (background), where it would otherwise be the figure (1978:108).6

Following this logic, Givón argues that negative contexts yield syntactic conservatism in cases of diachronic change that are aimed at “expressive-elaborative change, motivated by the creative drive to elaborate more complex and subtle nuances of meaning” (1979: 122). The innovation of a new construction ([tener que + Infinitive]) in a system that already includes two existing obligation constructions can arguably be considered this type of change. Therefore, one might expect negative contexts to disfavor the occurrence of tener que, instead favoring the older (conservative) forms.

This argument connects with the discussion of subjectivity in 1.2.1.1 through the notion of grounding. Following Langacker (1990), Traugott contrasts objective speech events with more subjective events:

[…] some are expressed in a maximally objective way, in the sense that events and their participants are placed ‘on-stage’ as the focus of attention, while elements of the ‘ground’ (the speech event itself, and its participants) are unprofiled, unexpressed, ‘off-stage’ as audience looking on. Other situations are expressed in a way that aligns some aspect of the event and its participants with the ground, and hence with some part of the viewing scene. (1995: 32-33)

Thompson (1998) provides another perspective on this issue in her discussion of the syntactic differences between negative and interrogative constructions by considering the interactive functions of the two. She argues that although their functions are similar, they are nonetheless distinct, and this distinction has consequences for their respective structures.
Based on Givón’s statement about the grounding effect of negation, it would appear that negation, grounding, subjectivity, and potentially syntactic conservatism are all correlated. Indeed, there is evidence to support this: Torres Cacoullos (2012) and Aaron (2010) both found that negation disfavored the occurrence of innovative Spanish forms in their analyses of the periphrastic present progressive and future constructions, respectively.\(^7\)

### 1.2.3. Aspect

Continuing the line of reasoning regarding subjectivity and figure/ground relationships, a third hypothesis can be generated, this time concerning aspect. Hopper (1982) states that a primary function of aspectual distinctions is to distinguish events from non-events. Aspect is utilized as a means of ordering events sequentially, effectively foregrounding certain events that are part of the sequence, relative to (back)grounded information that is not part of that sequence. The example in (13) illustrates this contrast.

\[(13) \quad \text{La emoción la ahogaba, y } \text{tuvo que sentarse sobre una piedra.}\]

‘The feeling was suffocating her, and she had to sit down on a rock.’
[19\(^\text{th}\) c.; Davies 2002-]

In (13) the feeling of being suffocated is an ongoing condition, whereas *tuvo que sentarse* ‘she had to sit down’ is rendered in the perfective Preterit, which bounds the event of sitting as a punctual occurrence. The punctual (perfective) event therefore contrasts with the grounded information (the emotional feeling of suffocation), placing the former in the foreground (Bybee, Perkins and Pagliuca 1994: 301).

\(^7\) It is worth noting, however, that there is no consensus on the reason why negation yields syntactic conservatism. Instead of Givón’s counter-expectation and grounding explanation, Torres Cacoullos suggests that negative polarity neutralizes tense-aspect distinctions (2012:46).
In her study of the historical synthetic pluperfect indicative form in Medieval Spanish, Klein-Andreu (1991) finds that the older form is frequently replaced by an innovative periphrastic form in contexts where an event is foregrounded. The older synthetic form is preferred for backgrounding events. This evidence can be applied to the present study to predict that the innovative grammaticalizing tener que construction will be favored in foregrounded contexts (i.e. with perfective aspect as in (13)) and disfavored in backgrounded contexts.

Testing the hypotheses presented here will be crucial to analyzing the variation between the grammaticalizing [tener que + Infinitive] and the alternative deontic modal constructions, especially its predecessors [haber de + Infinitive] and [deber (de) + Infinitive]. What is more, the present research will actually depend on the simultaneously discrete and continuous nature of the variables being examined. This is particularly true when analyzing the variation between [tener que + Infinitive] and the three other modal constructions.

1.3. A crosslinguistic tendency: possession > obligation

The predictions made above are aimed at accounting for the grammaticalization of a specific construction in a single language. But the hypotheses are based on tendencies in diachronic change that have been observed in various language families. In recent years, several authors have approached grammaticalization as a phenomenon whose manifestations exhibit consistent, recurring trends. Researchers like Heine and Kuteva (2002) and Traugott and Dasher (2002) have taken a somewhat taxonomic approach to presenting crosslinguistic tendencies in grammaticalization, focusing especially on the recurring types of semantic change. Bybee, Perkins and Pagliuca (1994) present a sample of genetically diverse languages, reporting the source forms from which markers of tense, aspect, and modality have been derived in the sample.
In each of these works, possession is listed as a common semantic source for forms that acquire deontic modal (obligation) meaning. Heine and Kuteva note that the development (or cline) from possession to obligation has been observed in Germanic, Romance, and Niger-Congo languages (2002: 243-254). Denning (1987) provides a more extensive list of languages exhibiting this development, including Celtic, Baltic, Slavic, and Nilotic examples. One of the aims of the present study is to place the specific example of the \([tener que + \text{Infinitive}]\) construction in the context of this crosslinguistic path of grammaticalization and observe the specific quantifiable elements that are operative in its evolution.

1.3.1. Previous studies

A number of studies have targeted the diachronic trajectories and the variation between modern modal forms in English (Brinton 1991; Fischer 1994; Nordlinger and Traugott 1997; Dollinger 2006; Tagliamonte 2004; Tagliamonte and Smith 2006). However, the historical development of Spanish modals has not been investigated so extensively. Yllera (1980) provides a thorough inventory of Medieval Spanish periphrastic constructions, including brief accounts of the possession-derived forms mentioned above \((tener que, haber de, haber que, deber (de))\). But her description provides little more than an acknowledgment of a construction’s presence in the language and a mention of its earliest appearance, yielding not much beyond a few paragraphs and a handful of examples for each.

Lacking a detailed account of the origin and grammaticalization of \([tener que + \text{Infinitive}]\), the evidence from its English equivalent provides something of a hypothetical analog for the research undertaken here. Brinton (1991) and Fischer (1994), synthesizing earlier work (van der Gaaf 1931; Visser 1969; Mitchell 1985), offer three basic stages of semantic and structural changes to account for the evolution of \(have to\) in English. Although imprecise, these
stages allow for a comparison of the findings presented in Chapters 2-5 with the predictions implied by the phases in Figure 1-1.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td><em>have</em> in the construction <em>have</em> + OBJECT + INFINITIVE is a full lexical verb meaning ‘possess’, while the nominal object denotes what is possessed and the infinitive serves as an ‘adverbial adjunct of purpose’, as in <em>I have food to eat.</em></td>
</tr>
<tr>
<td>Stage 2</td>
<td>The meaning of <em>have</em> changes, with possession and obligation meanings existing simultaneously. Syntactically, the structure is unchanged, but the object is now viewed as a responsibility or duty to be accomplished by the subject of <em>have</em>, as in <em>I have work to do.</em></td>
</tr>
<tr>
<td>Stage 3</td>
<td>Semantically, <em>have</em> now expresses obligation exclusively, and the object and infinitive have been transposed (<em>have</em> + INFINITIVE + OBJECT). The nominal object is now the object of the infinitive rather than <em>have</em> as in <em>I have to do work.</em></td>
</tr>
</tbody>
</table>

Figure 1-1. Proposed stages of semantic and structural evolution of English *have to*

Yllera alludes to a usage of *tener* with the relative pronoun *que* followed by an infinitive as the origin of the [*tener que* + Infinitive] construction, much like what is suggested in Stage 1 for *have to*. This is the type of arrangement found in early tokens like example (5) above: *et non tenemos ninguna cosa que comer* ‘and we do not have anything to eat’, where *que* designates the food as the object upon which the infinitive, or ‘adverbial adjunct of purpose’, is to be performed.

Still, Yllera provides no description of the intermediate changes that presumably took place as *tener que* transitioned from the first stage to its modern modal sense. In the following chapters, the evolution of the [*tener que* + Infinitive] construction is systematically traced from its earliest appearance through its acquisition of modal meaning. This is done to fill in the sizable gaps in Yllera’s limited description of *tener que* as well as compare its grammaticalization to the stages that have been proposed for its English equivalent *have to.*
1.4. The present study: design and methods

Motivated by the need for a thoroughly detailed account of the grammaticalization of 
[tener que + Infinitive] supported by methodical quantitative analyses, the research presented 
here examines data from several centuries of written and spoken (for the 20th century) sources. 
These include literary texts and theatrical plays as well as transcribed interviews; a full listing of 
the corpus is found in the Appendix.

Along with elucidating the history of the targeted construction itself, this study tests 
predictions made by the literature on grammaticalization as a general phenomenon of linguistic 
change. Although suggestive and intuitively appealing, many of the hypotheses discussed in 
Sections 1.1 and 1.2 have been left unexamined (if not altogether unsubstantiated). There is a 
small body of work by researchers that have begun to scrutinize grammaticalization using 
 systematic quantitative analyses, to which this investigation seeks to contribute in testing those 
hypotheses (Aaron and Torres Cacoullos 2005; Torres Cacoullos 2006; 2009; 2012; Poplack and 
Malvar 2007; Torres Cacoullos and Schwenter 2008; Bybee and Torres Cacoullos 2009; Copple 
2009; Aaron 2010, inter alios).

The present research is not intended as a comprehensive “audit” of the theoretical 
underpinnings of grammaticalization; rather, it focuses on a specific path of grammaticalization. 
The path from possession to obligation has been attested in numerous languages and can therefore 
be utilized as a case study to determine which of the presumed components of grammaticalization 
are identifiable via measurable evidence, and what their quantifiable manifestations are. The 
project is outlined in the following sections, divided into four chronological periods.
1.4.1. Chapter 2: Origins of the \textit{tener que + Infinitive} construction: 12\textsuperscript{th}-14\textsuperscript{th} centuries

The earliest data examined here is drawn from the 12\textsuperscript{th} century, effectively the earliest point at which the language can be considered Spanish rather than an ambiguous variety of Hispano-Romance. Using a corpus of written documents, it is possible to establish that the first appearance of the \textit{tener que + Infinitive} construction is within this epoch. The construction appears only a few times in these first three centuries of the analysis, and when it does it clearly conveys possession. Also, the relative clause component included in the construction typically features one of only a few lexical infinitives (e.g. \textit{comer} ‘to eat’, \textit{dar} ‘to give’).

\begin{equation}
\text{et non tenemos ninguna cosa que comer}
\end{equation}

\textit{‘and we do not have anything to eat’} [13\textsuperscript{th} c.; \textit{Calila}: 326]

Not only is the targeted construction exceedingly rare, but the verb \textit{tener} itself has not yet even established itself as the predominant form used to encode possession by this point. Instead, \textit{haber} ‘to have’ is clearly the preferred possession verb during this period, occupying several domains in which it would eventually be replaced by \textit{tener} (see especially Sections 2.1, 2.4.1, and 3.2). The latter verb illustrates evidence of its etymological meaning ‘to hold, to grasp’, and is commonly used to express a belief (a metaphorical sense of mentally grasping) in examples like that in (15).

\begin{equation}
aoruauan cada uno a quál-querie de-las planetas teniendo que aquel era su dios.
\end{equation}

\textit{‘Every one of them adored one planet or the other, believing that it was his god.’} [13\textsuperscript{th} c.; \textit{GE1}: I, 123]
Examples like these provide evidence of how the construction grammaticalizes, illustrating that *tener* acquires modal meaning as part of the specific construction in (14), instead of through the mere cooccurrence of the forms *tener* and *que* (15).

1.4.2. Chapter 3: Semantic and structural changes in [*tener que* + Infinitive]: the 15th-18th centuries

The second time period begins with the 15th century and ends with the 18th, capturing a crucial turning point in the grammaticalization of [*tener que* + Infinitive]. The verb *tener* ‘to have’ enjoys a substantial increase in its frequency of occurrence as it supplants *haber* in the domain of possession. Concomitant with this, the *tener que* construction also becomes more frequent, and is generalized to a steadily increasing number of distinct infinitive types. This is a break from the first time period, where the construction appeared only with a small group of lexical types.

In addition to its ever-increasing rate of occurrence, [*tener que* + Infinitive] exhibits clear semantic changes during this period. One of the lexical infinitive types that appear commonly in the construction is *hacer* ‘to do’, lending a sense of responsibility and/or intention to the overall meaning, as in (16). The specific construction [*tener que hacer*] ‘to have (something) to do’ attains a high enough frequency itself that it becomes a conventionalized instance of the general [*tener que* + Infinitive], illustrated in (17). Examples that indicate obligation more directly are found to follow thereafter (18).

(16)  *como si no tuviera otra cosa que hacer*

‘as if (he) did not have anything else to do’

[16th c.; Davies 2002-]
Verdad es, pero mal conoces a Celestina. Cuando ella **tiene que hacer**, no se acuerda de Dios ni cura de santidades.

‘That is true, but you know Celestina poorly. When she **has (something) to do**, she worries neither about God nor sanctity.’ [15th c.; *Celestina*: 222]

**no tenía que hacer la proposición** [...] ni aparecía proceso, testigos ni competencias algunas y sólo instaba la porfia de los acusadores insolentes

‘[he] did not **have to make a statement** [...] nor was there process, nor witnesses, nor arguments whatsoever, and he only encouraged the complaints of the insolent accusers’ [17th c.; Davies 2002-]

Accompanying this semantic change, the construction diverges from the structural arrangement that was characteristic of early tokens encoding possession. That is, fewer and fewer tokens are found with a pronominal object intervening between *tener* and the infinitive (i.e. *que* ‘something’), as in (17). Instead, the trend is toward tokens with postposed objects (18) and tokens with no overt object at all.

By the 18th century, [*tener que* + Infinitive] has clearly acquired modal meaning (see Section 3.2). However, it has not yet achieved a frequency of occurrence high enough to rival those of the older [*haber de* + Infinitive] or [*deber (de)* + Infinitive]. The *haber de* construction is ten times as frequent as [*tener que* + Infinitive], while the *deber (de)* is four times as frequent in the final century of this time period.

**1.4.3. Chapter 4: Variation between [*tener que* + Infinitive] and other periphrastic modal constructions: the 19th century**

In the 19th century, [*tener que* + Infinitive] has firmly established itself as part of the system of periphrastic modal constructions analyzed here. When viewed alongside the three alternative forms, [*tener que* + Infinitive] makes up 22% of the combined data (up from just 8% in the 18th century). The *tener que* construction now occurs 427 times per million words in the
19th century corpus, up from just 139 tokens per million in the 18th century. It also exhibits greater productivity, appearing with more than 200 distinct lexical infinitive types.

Given its substantial increase in frequency (both absolute and relative to the other three constructions), the tener que construction is examined using variable rule analysis (Sankoff 1988a) to investigate the variation between the targeted construction and the three others. The results of these analyses confirm several of the hypotheses discussed above in Sections 1.1 and 1.2. First person subjects do, in fact, favor the choice of [tener que + Infinitive] over the alternative forms, as does perfective aspect, supporting the prediction that foregrounded contexts favor the innovative construction (Section 4.7). Negative polarity disfavors its occurrence (§4.7), implying that negative contexts are syntactically conservative, per Givón’s (1978) suggestion.

1.4.4. Chapter 5: Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 20th century

The 20th century data continue the trajectory found in the transition from the 18th to the 19th century. The tener que construction makes up an even larger portion of the combined tokens of the four modal constructions, now accounting for 53% of the 20th century data by itself. Having made another large leap in frequency (up from 22% in the 19th century), the consistently upward trajectory of [tener que + Infinitive] illustrates the gradualness associated with grammaticalization. Still, gradualness does not entail a constant rate of change, as the particular grammaticalizing form investigated here was at first slow to gain frequency but accelerated in the final few centuries of the study. The older [haber de + Infinitive] has receded as quickly as [tener que + Infinitive] has gained ground: the former now accounts for only 3% of the combined data, down from over 60% just two centuries prior.
Multivariate analyses of the variation between [tener que + Infinitive] and the three other constructions were again utilized for comparison to the 19th century results. 1st person subjects and perfective aspect again favored tener que. Negative polarity disfavored its occurrence, but was not the most influential factor as it had been in the 19th century. This suggests that [tener que + Infinitive] may no longer be altogether innovative in the system as a whole. That is, the syntactic conservatism associated with negation may not have as strong an effect now that the incumbent tener que construction is the most frequent of the four.

This evidence, combined with that of the 19th century analyses, supports several of the predictions described above, regarding the contexts that were expected to favor the innovative tener que construction. At the same time, they provide detailed insight into the nature of the variation among the four deontic modal verb constructions included in the obligation system of Modern Spanish. In Chapter 6, the implications of these findings are discussed, compiling the results of the present research and synthesizing their contribution to both grammaticalization theory and our understanding of the specific path of development from possession to obligation.
Chapter 2

Origins of the [tener que + Infinitive] construction: 12th-14th centuries

The search for the origins of [tener que + Infinitive] ‘to have to + Infinitive’ necessarily begins with an investigation of some of the earliest documents written in Medieval Spanish. Written materials prior to the 12th century are limited, but the period from the 12th to the 14th century is marked by advances in standardization, especially under the direction of King Alfonso X, in the mid-to-late 13th century (Penny 2002: 20). What is more, as will be addressed here, the verb tener ‘to have’ does not emerge as a form encoding general possession in its own right until this period. Instead, the earlier verb haber ‘to have’ persists as the usual form used to encode possession through the 14th century (Corominas 1980). For these reasons, it would likely be fruitless to seek the formal and semantic sources of the modern [tener que + Infinitive] construction in writings any older than the 12th century.

In her inventory of periphrastic constructions in the history of Spanish, Yllera states that [tener que + Infinitive] is attested as early as the 13th century (1980: 117). She also describes a number of related constructions which include tener, along with marking their earliest occurrences. However, her description of the [tener que + Infinitive] construction is little more than an entry in a list, and lacks any sort of counts or frequencies for its use. The present investigation seeks to remedy this lack of quantitative detail by using the data found here to interpret the semantic status of the target construction and indeed the verb tener itself.

The study is arranged as follows: first, the semantic and functional status of the verb tener in the 12th-14th centuries is assessed, paying specific attention to the process by which it acquired its modern meaning ‘to possess’. Second, a descriptive account is provided of the various meanings, some etymologically conservative and others more innovative, to which tener is applied throughout the three centuries examined here. Where relevant, the contexts in which
tener appears will be compared to those of its competitor haber for the centuries in question. Finally, the source of modern [tener que + Infinitive] will be described, along with a presentation of tendencies that are evident in the earliest occurrences of the construction.

2.1. The development of possession meaning

Crucial to the development of the [tener que + Infinitive] construction are the semantic changes which tener underwent in its grammaticalization from a lexical verb denoting possession to a modal verb encoding obligation. Thus, an appraisal of the semantic status of lexical tener in this earliest time period is essential to outlining the processes at work in its gradual development of modal meaning and the subsequent changes in its function and frequency. Furthermore, the findings presented here will guide the subsequent investigation of the [tener que + Infinitive] construction as it grammaticalizes in later centuries.

Vaan defines teneō (tenēre), the Latin source of the Spanish infinitive tener, as ‘to hold, grasp’ (2008: 612). This is a common semantic source for markers of predicative possession, or Have-possession, as in John has a dog (Heine & Kuteva 2005: 290-291). This typological arrangement is described by Stassen (2011):

Among [the strategies for encoding possession], one stands apart in that it encodes the possessive relationship between possessor and possessed item in the form of a transitive construction. In this Have-Possessive, the possessor NP and the possessed NP function respectively as the subject and the direct object of a ‘have’-verb, which, in many cases, can be shown to derive from some verb indicating physical control or handling, such as ‘take’, ‘grasp’, ‘hold’, or ‘carry’.

The etymological source of Latin teneō is just this sort of meaning, drawn from the Indo-European root *ten-os- ‘to stretch’ (Vaan 2008: 613). The implication is that the original ‘stretch’
meaning has been augmented to take on a sense of stretching out the hand to take hold of or seize an object. From there, the punctual act of taking hold of an item can be given duration, yielding the more general ‘hold’ meaning of *teneō*, a sense that persists in various English verbs exhibiting –*tain*, the modern reflex of the Latin root (cf. *obtain, retain, detain*). Of particular importance to the present investigation, numerous occurrences of *tener* in the Old Spanish period evidently retain elements of the source meanings ‘seize’, ‘hold’ and ‘occupy’.

(19)  *e uio cuemo los d-affrica tenien el puerto*
and see.PST.3SG how those from-Africa hold.IPFV.3PL the port
*preso con su flota*
captive with their fleet
‘and (he) saw how those from Africa held the port captive with their fleet’
[13th c.; EE1: fol. 16r]

(20)  *et saltó el león en el asno de-trás por lo tener*
and jump.PST.3SG the lion on the ass from-behind for it capture.INF
‘and the lion jumped on the ass from behind in order to capture it’
[13th c.; Calila: 260]

(21)  *fue su señor a caça e salió un conejo, perdiol' e non l' pudo tener.*
‘His lord went on a hunt and out jumped a rabbit; he lost it and could not capture it.’ [14th c.; LBA: verse 1359]
Heine characterizes this change based on an ‘action schema’ involving an agent, a patient, and an action, i.e. ‘X takes Y > X has, owns Y’ (1997:47). This is a diachronic model, meaning that the ‘seize/take’ sense necessarily precedes the ‘hold’ sense. According to Givón (1984:103), the innovative ‘have’ meaning

...arises out of the semantic bleaching of active possession verbs such as ‘get’, ‘grab’, ‘seize’, ‘take’, ‘obtain’ etc., whereby the sense of “acting to take possession” has been bleached, leaving behind only its implied result of “having possession”.

Possession would presumably be limited at first to relatively small, tangible objects and then generalized to include larger objects (I have a house) and even abstract concepts (I have an idea). The uses of tener in (19-21) demonstrate the retention of earlier semantic components traceable to Latin tenēre. However, such examples occur alongside uses of tener that appear to indicate more general possession. These coexist with uses of haber (from Latin habeō ‘I hold, keep, have’), a form which predates tener in the expression of possession, having undergone the semantic transition from ‘take’ to ‘have’ in Latin (Baldi and Nuti 2009: 266).

(22) un rey muy poderoso, que avía nonbre
    a king very powerful, REL have.IPFV.3SG name

Varamunt, tenía una ave que dezían Catra
    Varamunt, have.IPFV.3SG a bird REL say.IPFV.3PL Catra

‘...a very powerful king, who had the name Varamunt, had a bird that they called Catra’ [13th c.; Calila: 272]
The possession encoded by *haber* in *avía nombre* ‘had the name’ is arguably distinct from the possession expressed by *tener* in *tenía una ave* ‘had a bird’. Nonetheless, the coexistence of the two forms complicates the task of determining the degree to which *tener* serves as a general possession verb in Old Spanish.

In order to arrive at a thorough description of *tener* and its possible functions during this period, a selection of texts was compiled, from which all occurrences of *tener* were counted and classified based on their semantic interpretation. Table 2-1 displays the distribution of documents selected for the analysis and the total number of *tener* tokens from each. All works are in prose with the exception of the *Cantar de mio Cid* and the *Libro de buen amor*, both in verse. Complete bibliographical information for the corpus used for this and the following chapters can be found in the Appendix.

<table>
<thead>
<tr>
<th>12th century</th>
<th>Words</th>
<th>tener tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cantar de mio Cid</em></td>
<td>35,500</td>
<td>80</td>
</tr>
<tr>
<td><strong>Century total</strong></td>
<td><strong>35,500</strong></td>
<td><strong>80</strong></td>
</tr>
<tr>
<td>13th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>General estoria. Primera parte</em></td>
<td>577,300</td>
<td>922</td>
</tr>
<tr>
<td><em>Calila e Dimna</em></td>
<td>87,200</td>
<td>217</td>
</tr>
<tr>
<td><em>Estoria de España. Primera parte</em></td>
<td>279,500</td>
<td>722</td>
</tr>
<tr>
<td><strong>Century total</strong></td>
<td><strong>944,000</strong></td>
<td><strong>1861</strong></td>
</tr>
<tr>
<td>14th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>El libro del caballero Zifar</em></td>
<td>34,000</td>
<td>116</td>
</tr>
<tr>
<td><em>Libro de buen amor</em></td>
<td>58,000</td>
<td>241</td>
</tr>
<tr>
<td><em>El conde Lucanor</em></td>
<td>79,200</td>
<td>288</td>
</tr>
<tr>
<td><strong>Century total</strong></td>
<td><strong>171,200</strong></td>
<td><strong>645</strong></td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td><strong>1,150,700</strong></td>
<td><strong>2586</strong></td>
</tr>
</tbody>
</table>

Tokens of *tener* were classified as belonging to one of two main semantic categories, each of which included multiple sub-categories. The source ‘to hold’ and innovative ‘to possess’ uses combine to make up the first category, while the more abstract meanings ‘to believe’ and ‘to
esteem’ comprise the second. Each category will be treated in turn, beginning with the basic lexical meanings.

2.2. ‘Hold, possess’-type uses of tener

For the seven documents included here, tener demonstrates several uses which are subsumed under the general category of literal ‘hold’ or ‘possess’ meanings. These two general senses, despite representing distinct stages in the semantic development of tener as a possession verb, are found here to include a number of nuanced and overlapping meanings.

2.2.1. ‘Hold’ vs. ‘possess’

Of the various meanings that are included here for tener, 1449 (56%) were classified as general holding or possession. However, only a few tokens can be interpreted as unequivocally expressing the older concrete sense of grasping an object in the hand.

(23) Moysen diziendo estas palabras tener el cayado en la mano como pastor.
Moses say.GER these words have.IPFV.3SG the crook in the hand like shepherd.

‘Moses, saying these words, held the crook in his hand like a shepherd.’
[13th c.; GEI: II, 81]

(24) et tener elle angel en la mano una espada sacada.
‘and the angel held in its hand a drawn sword’ [13th c.; GEI: II, 779]
Tokens for which there is a reference to *la mano* ‘the hand’ encode more overtly the concrete sense of holding an object, but these are uncommon. And even these occurrences do not entirely preclude an interpretation of more generalized possession rather than specific holding (i.e. ‘to have in the hand’ vs. ‘to hold in the hand’).

Presumably, this is precisely what allows for the pragmatic inference necessary for *tener* to add that semantic component to its meaning. From this perspective, which is thoroughly discussed by Hopper and Traugott (1993: Chapter 4), the context in which a construction occurs can supply an additional semantic component to the form. If this additional element of meaning becomes conventionalized, it is effectively added to the inherent semantic content of the form itself, regardless of context. This is a slightly different perspective on meaning change in grammaticalization than that of Givón (1984) discussed in Section 2.1, which emphasizes bleaching, or loss, of semantic content. Still, the two are not necessarily mutually exclusive; as mentioned in Section 1.1.1, forms undergoing semantic change (especially in grammaticalization) evidently experience both the addition and loss of certain semantic elements.

The tokens in (25-26) exemplify the potential for pragmatic inference and subsequent generalization of the possession meaning⁸. The more frequent usage of *tener* is at least somewhat ambiguous:

(25) *Dos espadas tenedes* *fuertes* et *taiadores*

*two swords have.PRES.2SG strong and slicing*

‘you have two swords, strong and slicing’ [12th c.; *Cid: verse 2726*]

---

⁸ Pragmatic inference is discussed in more detail in Section 3.2.
Without an explicit mention of the hand, *tenedes* is just as readily interpreted as ‘you have’ as it is ‘you hold’, despite referring to a hand-held weapon in (25). Given the mention of a quantity of money owed to the speaker in (26), the reading here is one of general possession: the speaker has been paid a portion of the debt, and therefore possesses that quantity.

An additional portion of the tokens for which *tener* encodes a ‘hold’ meaning are somewhat more abstract. A number of tokens in the *General Estoria: Primera Parte* include references to divine covenants or laws, leading to an interpretation of *tener* that would be close to ‘keep (as a promise)’ or ‘uphold’. These meanings are extended away from the core senses of ‘to hold’ or ‘to have’ in that they refer to abstract objects, but they are considered here to belong to the more conservative of the two main categories of *tener* tokens. Although the object of a ‘keep’ token is abstract (e.g. a covenant, a custom), the conceptual link to the source ‘to hold’ is fairly transparent. Examples (27-29) demonstrate this usage in the context of an agreement or law.

(27) *Delos bienes que nuestro sennor prometio alos que*  
of-the goods REL our lord promise.PST.1SG to-DEF.PL REL  
*los sus mandados bien touiessen.*  
DEF.PL his commands well *keep.PST.PRF.3PL*  
‘Of the goods that our Lord promised to those who had kept his commands well’  
[13th c.; *GE1: II, 610*]
(28) **ellos que dexen las costumbres et la ley que**

they REL leave.SBJV.3PL the customs and the law REL

**sus padres touieron**

their parents keep. PST.3PL

‘those who leave the customs and the law that their parents upheld’

[13th c.; GE1: II, 792]

(29) **aquel prometimiento tenudo fuese et complido.**

DEM promise keep.PTCP be.PST.PR.F.3SG and fulfill.PTCP

‘That promise was kept and fulfilled.’ [13th c.; GE1: II, 613]

2.2.2. Competition between *tener* and *haber*

Due to its evident application to both the etymological source meaning (‘to hold’) and the innovative generalized meaning (‘to have’), it would be premature to conclude that *tener* has firmly established itself as a possession verb. Considering Givón’s (1984) aforementioned description of the semantic change in progress, it appears that the ‘have’ component has been added (or has begun to be added), but the ‘hold’ component has not yet been bleached from its overall meaning.

Further evidence of the incomplete acquisition of possession meaning by *tener* comes from the frequent use of *haber* for that function. Uses such as those in (30-31) persist through the 14th century.
comença a desviar se de la paz de los Romanos de

begin.PST.3SG to diverge REFLEX from the peace of the Romans for

non la querer tener ni auer-la con ellos.

NEG it want.INF have.INF nor have.INF-it with them

‘he began to distance himself from peace with the Romans for not wanting to have (tener) nor keep (haber) it with them.’ [13th c.; EEI: fol. 145v]

Si pudiessemos matar el moro Avengalvon, quanta riquiza

If be_able.PST.PRF.1PL kill.INF the moor Avengalvon as_much wealth

tiene auer la yemos nos.

have.PRS.3SG have.INF it FUT we

‘If we were able to kill the moor Avengalvon, what wealth he has (tener), we will take (haber).’ [12th c.; Cid: 2662]

During this period, tener and haber overlap in this semantic domain. Haber, the older of the two forms used to encode possession, has begun to grammaticalize further. In doing so, it takes on additional applications not available to tener. As found in the Cantar de mio Cid, haber has already begun to acquire its more modal and grammatical functions (prior to subsequently losing nearly all of its lexical meaning in Modern Spanish). These functions include that of the future construction in (13), auerla yemos nos ‘we will have it’. In this construction, haber is evidently being used to refer to the future when serving as an auxiliary verb (yemos ‘we will’), while also maintaining its basic possession meaning when used as a lexical verb (auerla ‘to have it’).9

Nevertheless, the two forms remain as viable alternatives for encoding the notion of general possession. Confirmation of this ongoing competition is available in several sources.

---

9 Fleischman (1982) provides a detailed account of the process by which habeō (haber) developed its meaning and function as an auxiliary verb to refer to the future in Romance languages.
Penny (2002: 194) asserts that tener did not take over the lexical (possession) role of haber until the 15th century. Glossaries found in critical editions of the texts utilized for the present analysis also allude to this overlap. In the Delgado Campos (1936) edition of the Libro de Buen Amor, the glossary entry for ey, an orthographic variant of he (from haber), lists both he ‘I have’ and tengo ‘I have’ (from tener) as definitions. Menéndez Pidal (1969), in his volume dedicated to the lexicon of the Cantar de mio Cid, also defines both auer (haber) and tener as ‘to possess’.10

Another domain in which tener has not yet supplanted haber is that of expressions of emotion and physical sensations. In Modern Spanish a number of such expressions include tener: tener miedo ‘to be afraid’ (literally ‘to have fear’); tener hambre ‘to be hungry’; tener sed ‘to be thirsty’; tener frío ‘to be cold’. But in the period analyzed here, these typically occur with haber.

(32) et ovo tan grand miedo, que non sabía si
and have.PST.3SG such great fear, COMP NEG know.IPFV.3SG if
era muerta o biva
be.IPFV.3SG dead or alive
‘and he had (haber) such great fear that he did not know if he was dead or alive’
[14th c.; Lucanor: 191]

(33) si a fanbre et non le dan a comer, et
if have.PRS.3SG hunger and NEG him give.PRS.3PL to eat.INF and
si a sed et non le dan a bever
if have.PRS.3SG thirst and NEG him give.PRS.3PL to drink.INF
‘if he has (haber) hunger and they do not give him to eat, and if he has (haber)
thirst and they do not give him to drink’[13th c.; Calila: 117]

---

10 A thorough evaluation of the semantic and functional overlap between haber and tener in the history of Spanish will be avoided here. However, several authors have written explicitly on the subject (cf. Seifert 1930; Gallarín 2008).
This is true for all of the texts included in the analysis, but in the 14th century Libro de buen amor, the innovative tener appears to be encroaching on haber in this domain. In fact, the document includes examples of both haber miedo and tener miedo (35-36) in addition to one example of tener frío (37):

(35) non an de Dios miedo, nin de sus amenaças

‘Do not have (haber) fear of God, nor of his threats.’ [14th c.; LBA: verse 415]

(36) que miedo vano tenemos

‘such vain fear we have (tener)’ [14th c.; LBA: verse 1447]

(37) Yo l’ dix: «Frío tengo, e por eso vengo a vos»

‘I told him “I am cold, and for that I come to you”’ [14th c.; LBA: verse 1026]

Once again, haber appears to retain its position in the realm of general possession. However, tener appears to be gradually acquiring this role and appearing in several of the contexts associated with its modern distribution. Thus, the concept of possession is entrenched in these constructions, rather than the lexical item haber, which is renovated with tener. Bybee describes this sort of conventionalized meaning as “packaging concepts for speaking” (2010:219; cf Slobin’s (1996) “thinking for speaking”)

(34) Quando has frío a los pies, ¿qué es lo que fazes?

‘When you have a chill in your feet, what is it that you do?’[13th c.; Calila: 353]
2.2.3. The possessive/resultative construction with tener

The possessive/resultative construction represents another subcategory of the ‘hold/possess’ tokens in which tener and haber have not yet achieved their respective modern distributions. Resultatives indicate that a present state exists as a result or consequence of an action performed in the past, and are the source for perfect constructions based on ‘have’ verbs in numerous European languages (Bybee, Perkins and Pagliuca 1994: 54). This is the case for English, in which have was utilized in a resultative construction that developed into the modern perfect construction, e.g. *I have the sculpture carved* vs. *I have carved the sculpture*. In the former case, possession of the object (*sculpture*) is the result of the action encoded by the participle *carved*. For Spanish, haber was utilized for a semantically parallel construction in which it also grammaticalized, yielding the modern perfect construction: *Yo he tallado la escultura* ‘I have carved the sculpture’\(^{11}\).

During the time period addressed here, haber appears in resultative constructions such as that in (38-39), as does tener (40-41). 198 (14%) of the 1449 ‘hold/possess’ tokens of tener occurred with a passive participle to form the possessive/resultative construction.

\textit{Haber:}

(38) \begin{align*}
\text{Por miedo de-} & \text{1 Rey Alfonso, que } \text{assi}
\end{align*}
For fear of-the king Alfonso, COMP thus

\begin{align*}
\text{lo } & \text{auien } \text{parado}
\end{align*}
OBJ.MASC.SG have.IPFV.3PL plan.PTCP.MASC.SG

‘For fear of King Alfonso, thus they had planned it’ [12\textsuperscript{th} c.; \textit{Cid}: verse 33]

\(^{11}\) For more complete descriptions of the process by which ‘have’ verbs grammaticalize in resultative constructions to yield perfects, see Dahl 1985; Bybee et al. 1994; Carey 1995.
(39)  
\textit{Andan los días et las noches et passada}

Walk.PRS.3PL ART days and ART nights and pass.PTCP.FEM.SG

\textit{han la sierra}

have.PRS.3PL ART mountain_range.FEM.SG

‘They march day and night and have passed the mountain range’

[12\textsuperscript{th} c.; \textit{Cid}: verse 1823]

\textit{Tener:}

(40)  
\textit{touo todo esto bien parado}

have.PST.3SG all DEM.MASC.SG well stand.PTCP.MASC.SG

et toda su tierra en paz

and all POSS land in peace

‘he had all of this in good stead and all his kingdom in peace’

[13\textsuperscript{th} c.; \textit{GE1}: I, 199]

(41)  
\textit{se non puede yr d-end assi como}

REFL NEG be_able.PRS.3SG go.INF from_within as like

si-l touiesse atado

if-OBJ.MASC.SG have.IPFV.SBJV.3SG tie.PTCP.MASC.SG

‘he cannot leave from there as if he were tied’ [13\textsuperscript{th} c.; \textit{GE1}: I, 318]

The \textit{haber} resultative diverges from the \textit{tener} resultative in that the former continued to grammaticalize into a perfect construction, whereas the latter did not\textsuperscript{12}. Evidence for the ongoing development in the \textit{haber} perfect is found in the lack of number and gender agreement between the erstwhile possessed object and the participle in that construction. Early examples of the

\textsuperscript{12} In standard Modern Spanish, the \textit{tener} resultative construction occurs with number and gender agreement between the participle and the object, with limited exceptions (Butt and Benjamin 2004: 226).
haber resultative exhibit agreement, as with (39) above, in which the participial adjective pasada
‘passed.FEM.SG’ agrees with the object sierra ‘mountain_range.FEM.SG’. Later examples of the
same grammaticalizing construction lack agreement, whereas it is maintained for the tener
resultative, which does not achieve the status of a perfect construction. Fruyt provides a thorough
account of the grammaticalization of Latin habeō which yielded, in this context, the modern

The examples in (42-43), which are drawn from the same document, indicate the
differing patterns of agreement for haber and tener: the example with haber (42) lacks
agreement, while the example with tener (43) does not. In (43), the participial adjective cercada
‘surround.PTCP.FEM.SG’ agrees with the object villa ‘town.FEM.SG’. The participle pasado
‘endured’ (with the invariable –o suffix which historically corresponded with the masculine
singular) in the haber construction in (24) does not agree with the object verguencas
‘embarassments.FEM.PL’.

(42)  non quisiera auer pasado las
      NEG want.IPFV.SBJV.3SG have.INF undergo.PTCP the
      verguenças
      embarrassment.FEM.PL
      ‘he would not have wanted to have suffered the embarrassments’
      [14th c.; Zifar 72]

(43)  su huesped le dixo de commo tenien
      POSS guest OBJ say.PST.3SG of how have.IPFV.3PL
      cercada la villa los sus enemigos.
      surround.PTCP.FEM.SG ART village.FEM.SG ART POSS enemies
      ‘his guest told him of how his enemies had the village surrounded’
      [14th c.; Zifar]
2.2.4. ‘Stay’

An additional application of the verb tener in the data analyzed here is a reflexive usage, sometimes with the reflexive pronoun se to mean ‘stay’ or ‘hold oneself’. With only 48 occurrences in the data examined here, this usage is less frequent than any of the other sub-categories of the ‘hold/possess’-type tokens. Tener(se) ‘to hold oneself, to stay’ is relatively literal and conservative in its meaning, in that it invokes the earlier ‘hold’ sense of teneō (tenēre). Therefore, such uses of tener appear to be among the more conservative in the corpus, as these tokens illustrate retention of some of the older elements of the form’s semantic content:

(44) *murieron muchos con el quándo quemo el fuego los dozientos et cinquánta uarones que tenien con chore*

‘Many died with him when the fire burned the 250 men that remained with Chore.’ [13\textsuperscript{th} c.; GE1: II, 807]

(45) *pero que se uos tenga luengo tiempo. nol taiades los aruoles que lieuan fructo de comer.*

‘But if you stay for a long time, do not chop the trees that bear fruit.’

[13\textsuperscript{th} c.; GE1: II, 905]

(46) *Et andido con dios et touo se toda uia con el.*

‘And he walked with God and stayed always with him.’ [13\textsuperscript{th} c.; GE1: I, 48]

2.2.5. ‘Extend’

Perhaps the most conservative of all the types of tener tokens, uses of the verb to mean ‘to extend’ are limited to the 13\textsuperscript{th} century documents examined here. This facet of the semantic
content of tener evidently draws on the meaning of its source, the Indo European root *ten- ‘to stretch’. The verb tender (also from PIE *ten- ‘to stretch’) expresses this notion in the modern language, as it includes ‘to extend, to reach’ among its several different meanings.

When these ‘extend’ uses of tener are combined with the ‘hold/possess’ tokens, they make up 7% (104/1553) of the total conservative ‘hold/possess’-type tokens. However, all of these are drawn from two documents: 85 from the Estoria de España: Primera parte and 19 from the General Estoria: Primera parte. The majority of these uses relate to spatial expanses, especially surveys of the geographical boundaries of territories and kingdoms, as in (47).

(47) ...et comiença en Teuca et tiene dalli fastal río Eufrates
‘...and it begins in Teuca and extends from there to the River Euphrates’
[13th c.; GEI: II, 291]

(48) fizo de naves en dos meses una puente que tiene de calcedonia fasta costantinopla
‘In two months he made a bridge of ships that extended from Calcedonia to Constantinople.’ [13th c.; EE1: folio 138r]

A few examples exhibit an abstract application of the ‘extend’ meaning to the temporal domain, most frequently in the delineation of bloodlines and important epochs:

(49) Et touo el primero destos tres tiempos segund dizen unos del comienço de Adam fasta moysen

---

13 The ‘extend’ uses of tener were not included as part of the ‘hold/possess’-type tokens when those uses were discussed above because their meaning appears to be drawn directly from the earlier (i.e. more conservative) ‘stretch’ sense of the verb’s etymological source.
‘And the first of these ages extended, as some say, from the beginning of Adam until Moses.’ [13th c.; GEI: I, 184]

(50) *se comienza de luego que amanesce et tiene fasta que ennochesce*

‘It begins once the sun rises and lasts until it sets.’ [13th c.; GEI: II, 165]

That these spatial and temporal uses of *tener* only appear in two documents is likely a product of the genre of these two texts, as each is in the format of a chronological history; the *Estoria de España* recounts that of Spain while the *General Estoria* tells of the Bible’s Old Testament. The ‘extend’ tokens of *tener*, albeit limited in number, illustrate that while *tener* is apparently in the process of acquiring new meanings and being extended to numerous semantic domains during this period, it nonetheless retains a telling link to its etymological source.

### 2.3. *tener* for expressing beliefs and mental assessments

Numerous tokens of *tener* in the corpus exhibit abstract metaphorical applications of the verb’s basic ‘to hold’ meaning. These occurrences encode several slightly different notions, but all have ‘to believe’ as their semantic center of gravity. In a general sense, the extension of *tener* from ‘to hold’ to ‘to believe’ is parallel to that of English *tenet* ‘a thing that is held (as true)’.

1133 (44%) of the 2586 total tokens of *tener* fall into this category. Evidence of this extension of the literal ‘to hold, to have’ meaning can be found in examples like that in (51). This type of usage is classified here as a lexical ‘hold, possess’-type token, but the context drives an inference that the implicit possessed object is abstract (e.g. a thought, idea or belief):
Et por lo que yo tengo en el corazón conosco lo que tú tienes en el tuyo.

‘And because of what I have in my heart, I know what you have in yours.’

[13th c.; Calila: 276]

Tokens including a variation of tener en el corazón ‘to have in one’s heart’ are seemingly conventionalized and persistent, as they occur in all of the documents considered here. In their structure, they parallel the concrete tener en la mano ‘to hold in the hand’ tokens discussed above, but differ in that the object is commonly an emotion, or other abstract entity.

esto es quándo tiene ell omne a dios en todo

‘this is when a man has God in all his heart and his will and his soul’

[13thc.; GE1: II, 253]

tenia en mi coraçon de ser vençidos aquellos caualleros

‘I had in my heart that those knights were vanquished.’ [14th c.; Zifar]

While referring to abstract possessed objects, tokens such as those in (52-53) do not create an altogether distinct meaning for tener. The sense is still one of holding or possession, but of an

---

14 The structure of tener en el corazón ‘to have in one’s heart’ varies in that it sometimes occurs with an article (e.g. el ‘the’) as in (32) or a possessive pronoun (e.g. su ‘his/her’) as in (33-34). The example in (33) is arranged differently, lacking the preposition en ‘in’ and including a human, rather than abstract, object.
abstract object rather than a concrete one. They do, however, provide a conceptual bridge to two subcategories of tokens: instances in which tener can be interpreted to mean ‘to believe’ and others for which it is somewhat closer to ‘to esteem’ or ‘to regard as’.

2.3.1. ‘To believe’

A frequent usage of tener has the innovative but thoroughly established (in the corpus) meaning ‘to believe’. 463 tokens (18%) of the 2586 total are of this type, and are almost always accompanied by the complementizer que ‘that’, followed by a finite clause, as in (54-56). This yields a schematic [tener que X] ‘to believe that X’. In English, the literal translation to have it that (i.e. ‘to believe that’) encodes this sentiment, as does the interpretation of has when used in the idiomatic expression rumor has it. Tokens of [tener que X] ‘to believe that’ appear in all of the documents in the corpus and remain common in the 14th century.

(54) Assi lo tenerien las yentes que mal ferido es
Such OBJ have.IPFV.3PL ART people COMP badly wounded be.PRS.3SG de muert.
of death
‘So believed the people that he is wounded unto death.’ [12th c.; Cid: verse 3641]

(55) aorauan cada uno a cual-querie de-las planetas teniendo
adore.IPFV.3PL each one ACC which-ever of-the planets have.GER que aqueI era su dios.
COMP DEM be.IPFV.3SG POSS god
‘Every one of them adored one planet or the other, believing that it was his god.’
[13th c.; GE1: I, 123]
(56) El conde tovo que Patronio le decía la verdad

ART count have.PST.3SG COMP Patronio OBJ tell.IPFV.3SG ART truth

‘The count believed that Patronio was telling him the truth.’

[14th c.; Lucanor: 119]

There are additional examples that show tener alternating with creer ‘to believe’, the verb for which this meaning persists in Modern Spanish. In fact, in his Vocabulario Medieval Castellano, Cejador y Frauca lists only creer as the definition for tener (1968: 383). The examples in (57-58) illustrate the alternation between these two verbs, sometimes used simultaneously. This simultaneous usage suggests that perhaps there is a subtle difference in meaning, but given the evidence above, it is clear that tener is again intended to mean ‘to believe’.

(57) Ell Emperador mando las todas tener et creer tan bien cuemo los quatro euangeliros

‘And the emperor commanded that all trust (tener) and believe (creer) as well the four gospels.’ [13th c.; EE1: folio 117v]

(58) Mas es de saber alos fieles de xpisto et creer et tener lo por firme.

‘But it is to be known to the followers of Christ and to be believed (creer) and held (tener) as certain.’ [13th c.; GE1: II, 302]

2.3.2. ‘To esteem’, ‘to regard as’

In the other subcategory of mental assessment uses of tener, the verb has a slightly different meaning. The semantic distinction is accompanied by a formal difference in the construction in which it typically occurs: [tener por X], where the category X is a predicate
adjective or predicate nominal. In these uses, tener appears followed by the preposition por ‘for, as’, yielding the meaning ‘to esteem as’ or ‘to regard as’ (literally ‘to hold as’), a sense that has been preserved to a limited extent in the modern language. This type of usage is even more common than the [tener que X] ‘to believe that’ tokens, occurring 570 times in the corpus (comprising 22% of the data). Tener por is often followed by a noun, the metaphorical object of tener:

(59)  a-uos tiene por señor

ACC-you have.PRS.3SG as lord

‘He regards you as his lord.’ [12th c.; Cid: verse 1339]

(60)  fijo de Japetho el gigant fue fallado et tenudo por el mas sabio et mayor estrellero

‘the son of Japeth the giant was found and esteemed as the wisest and greatest astronomer.’ [13th c.; GE1: I, 545]

In other instances, tener por precedes an adjective, but with an identical meaning, and again with an object (whether human or inanimate):

(61)  Muchos tienen por Ricos los yfantes de Carrion.

‘Many esteem as wealthy the princes of Carrion.’ [12th c.; Cid: verse 2510]

(62)  don Iohan tovo este exiemplo por muy bueno

‘Don Juan regarded this example as very good.’ [14th c.; Lucanor: 114]

The frequency of the [tener por X] schema may be slightly inflated due to its repeated occurrence in one document, El Conde Lucanor. 92 of the 288 tokens of tener in this text are of
this type, as it gives the account of an advisor offering advice to his Count (don Iohan in (62)) through fables and didactic tales. At the end of each story, the count typically ‘finds the advice to be good’. Still, when the tener por ‘to regard as’ construction is considered alongside the tokens of [tener que X] ‘to believe that’, we can see that ‘believe’ and ‘esteem’ uses are highly frequent throughout the corpus and represent a sizable portion of the total occurrences of the verb tener from the 12th to the 14th century (Table 2-2).

<table>
<thead>
<tr>
<th>Century</th>
<th>12th</th>
<th>13th</th>
<th>14th</th>
</tr>
</thead>
<tbody>
<tr>
<td>tener por ‘to esteem as’, ‘to regard as’</td>
<td>11</td>
<td>434</td>
<td>125</td>
</tr>
<tr>
<td>tener que + X ‘to believe that X’</td>
<td>4</td>
<td>323</td>
<td>136</td>
</tr>
<tr>
<td>Total tener tokens</td>
<td>80</td>
<td>1859</td>
<td>643</td>
</tr>
</tbody>
</table>

| Total mental assessment uses/Total tener tokens (%) | 15/80 (19%) | 757/1859 (41%) | 261/643 (41%) |

It is especially striking that tener has been so consistently used to express beliefs and assessments of individuals or propositions given that tener is not yet the dominant possession verb; it continues to compete with haber during this time. These abstract metaphorical applications of its meaning have been added, seemingly by way of the more conservative and concrete ‘to hold’ sense, rather than via the notion of possession.

2.4. Origins of the [tener que + Infinitive] construction

The final subset of tener tokens remaining to be addressed is the target of the current project. In a small portion of the occurrences of tener for which the interpretation is one of concrete possession, the possession verb is followed by the relative pronoun que ‘which’ and an infinitive verb. Yllera states that the construction is already present in the language by the 13th
century (1980: 117). The present data corroborate this, as demonstrated by the example in (63), from the 13th century *General Estoria*.

(63)  nin *tenemos*  pan  *que*  comer  nin  otra  cosa  ninguna

neither have.PRS.1PL bread REL eat.INF nor other thing neither

‘We have neither bread (which) to eat nor any other thing.’ [13th c.; GE1: II, 179]

In *pan que comer* ‘bread to eat’, the relative pronoun serves to specify *pan* ‘bread’, effectively selecting the instance of bread that is possessed and designated to eat from the set including all possible bread. Tokens like that in (63) are semantically identical to the construction that has been described as the first stage of the grammaticalization path by which the English [*have to + Infinitive*] construction developed (Brinton 1991; Fischer 1994). Figure 2-1 displays the three basic stages for the development of [*have to + Infinitive*].

| Stage 1 | *have* in the construction *have +* OBJECT + INFINITIVE is a full lexical verb meaning ‘possess’, while the nominal object denotes what is possessed and the infinitive serves as an ‘adverbial adjunct of purpose’, as in *I have food to eat*.
| Stage 2 | The meaning of *have* changes, with possession and obligation meanings existing simultaneously. Syntactically, the structure is unchanged, but the object is now viewed as a responsibility or duty to be accomplished by the subject of *have*, as in *I have work to do*.
| Stage 3 | Semantically, *have* now expresses obligation exclusively, and the object and infinitive have been transposed (*have + INFINITIVE + OBJECT*). The nominal object is now the object of the infinitive rather than *have* as in *I have to do work*.

Figure 2-1. Developmental stages in grammaticalization of English *have to + INFINITIVE*

Presuming that the origin of the Spanish [*tener que + Infinitive*] lies within the Spanish equivalent of the Stage 1 construction in Figure 2-1, all tokens of [*tener + (X) + que + Infinitive*]

---

15 It is worth noting that even a relatively objective statement like *I have food to eat* is not entirely without a sense of intention or even responsibility, given the physiological need to eat.
were extracted from the corpus. Surprisingly, this extraction yielded only 14 occurrences of the target construction, meaning that it made up only 1% of the 1553 total tokens classified as ‘hold/possess’ uses of tener. Table 2-3 demonstrates the low frequency of the target construction with respect to the other applications of tener described above.

Table 2-3. Token counts for each use of tener by century.

<table>
<thead>
<tr>
<th>Type of use</th>
<th>12th C.</th>
<th>13th C.</th>
<th>14th C.</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘to hold’ / ‘to possess’</td>
<td>65</td>
<td>806</td>
<td>318</td>
<td>1189 (46%)</td>
</tr>
<tr>
<td>tener resultative</td>
<td>-</td>
<td>140</td>
<td>58</td>
<td>198 (8%)</td>
</tr>
<tr>
<td>‘to stay’</td>
<td>-</td>
<td>43</td>
<td>5</td>
<td>48 (2%)</td>
</tr>
<tr>
<td>‘to extend’</td>
<td>-</td>
<td>104</td>
<td>-</td>
<td>104 (4%)</td>
</tr>
<tr>
<td>Stage 1 [tener + (X) + que + Infinitive]</td>
<td>-</td>
<td>11</td>
<td>3</td>
<td>14 (1%)</td>
</tr>
<tr>
<td>‘to believe’</td>
<td>4</td>
<td>323</td>
<td>136</td>
<td>463 (18%)</td>
</tr>
<tr>
<td>‘to esteem’ / ‘to regard as’</td>
<td>11</td>
<td>434</td>
<td>125</td>
<td>570 (22%)</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>1861</td>
<td>645</td>
<td>2586</td>
</tr>
</tbody>
</table>

A comparison between English I have bread to eat and I have to eat bread illustrates the difference between this initial stage and the later, more grammaticalized construction. The former conveys possession of an object, bread, and qualifies it with to eat, which presumably also carries a sense of intention. The latter indicates an obligation to consume the bread, which is no longer the object of have but instead of to eat.

This semantic difference is also signaled by a change in the word order. In the construction which encodes obligation, have and to are contiguous, whereas the earlier (possession) construction follows the pattern have + object + to + verb. Dollinger calls this the “discontinuous” version of construction, as the object (e.g. bread) intervenes between have and to (2006: 292). It is in the contiguous context, where have and to are adjacent, that Fischer argues the English construction began down the grammaticalization path which transformed have to into a modal (or “semi-modal”) form. In her estimation, the crucial element is not simply the
appearance of an infinitive complement with the possession verb, but the juxtaposition of have and its infinitive complement (1994: 137).

Incidentally, tokens of each type (contiguous and discontinuous) are found in the present data. In several instances, a concrete object intervenes between tener and the relative pronoun que (64), while in others the object is implicit, leaving tener and que adjacent (65).

(64)  *dond tengo yo carnes que dar a tan mucha yent como esta*

‘Where do I have meat (which) to give to so many people as these?’
[13\textsuperscript{th} c.; *GE1*: II, 696]

(65)  *el que non tiene que dar, su caballo non corre*

‘he who does not have (anything) which to give, his horse will not carry him’
[14\textsuperscript{th} c.; *LBA*: verse 512]

It is difficult to determine whether the same can be said for Spanish [*tener que* + Infinitive], considering Fischer’s assertion that adjacency between have and to was crucial to the construction’s grammaticalization. In the scant amount of tokens found for the 12\textsuperscript{th} to 14\textsuperscript{th} centuries, 8 of 14 exhibit intervening material between tener and the relative pronoun que, as in (63-64). For six of these, the intervening material was an object, but in the other two, it was the adverb *ya* ‘anymore’:

(66)  *et dixieronle que njnguna cosa non tenien ya quel uender*

‘and they told him that they had nothing anymore to sell him.’
[13\textsuperscript{th} c.; *GE1*: I, 478]
Given the extremely low frequency of the target construction, it is difficult to make any generalizations about tendencies that appear in the tokens found here. To supplement the present data, tokens of the Stage 1 [tener + (X) + que + Infinitive] were extracted from the online Corpus del Español for the 12th to 14th centuries (Davies 2002-). This portion of Davies’ corpus comprises 9,746,974 words and includes most of the documents in the corpus used in the present study. Therefore, the supplemental tokens extracted were in addition to those already mentioned, in order to avoid counting any recurring tokens twice.

The procedure for extracting these supplemental tokens was not exhaustive, but nearly so, and involved the following: a search of the corpus from the 12th to 14th centuries was conducted, targeting all forms of the verb tener, where it co-occurred with an infinitive within the six following words. This allowed for capturing any intervening material between tener and the infinitive. For most of the tokens, this was nothing more than the relative pronoun que, as even explicit possessed objects were relatively rare. However, in a few tokens, there were as many as four intervening words, including que, the possessed object, and an indirect object pronoun (beneficiary):

(67) tiraron los frenos a-llos cauallos et dexaron los paçer
remove.PST.3PL ART bits to-the horses and let.PST.3PL OBJ graze.INF
ca non tenjan otra çeuada que les dar
because NEG have.IPFV.3PL other barley REL OBJ give.INF
‘They removed the bits from the horses and left them to graze because they did not have any other barley to give them.’

[14th c.; Davies 2002-]
The initial search yielded a list of all of the infinitives that occurred in the crucial context, as well as the number of times they occurred there. For example, the infinitive *dar* ‘to give’ appeared 484 times in the search context (*tener* followed by an infinitive within six words). Of these, 20 (4%) of the *dar* tokens were in the targeted construction [*tener* + (X) + *que* + Infinitive], making it the most common infinitive. Given this low frequency, and that only 1% of the possessive *tener* tokens analyzed in the present investigation occurred in the target construction, infinitives with counts fewer than 10 total appearances were not examined, as they were not expected to yield any tokens. In all, the search returned 40 tokens of the *tener que* construction for the 12th, 13th and 14th centuries.

Along with the 20 tokens of *dar* ‘to give’, there were 15 tokens of *comer* ‘to eat’, easily making them the two most frequent infinitives appearing in the [*tener* + (X) + *que* + Infinitive] construction in Davies’ corpus. Interestingly, the same can be said for the tokens extracted from the corpus utilized here. Of these 14 tokens, there were 7 with *comer* ‘to eat’ and 4 with *dar* ‘to give’. Thus, a suggestive tendency emerges: the infinitives *dar* and *comer* account for 44% and 41%, respectively, of the 54 combined tokens of the target construction.

These specific conventionalized strings, or prefabs, seem to act as a sort of “advance guard” for the general construction. The high frequency of occurrence of *tener* (X) *que dar* and *tener* (X) *que comer* presumably allows them to grammaticalize more quickly than the general [*tener* + (X) + *que* + Infinitive] construction and consequently contribute to its meaning. Thus, they lead the way in adding a sense of intention or even responsibility, given the physiological need to eat (*comer*) and the common association of giving (*dar*) with an exchange or debt (see Stages 1 and 2 in Figure 2-1 above). Furthermore, they facilitate the application of the construction to a broader range of lexical infinitive types (see Section 3.2). Bybee and Torres Cacoullos (2009) find evidence for this interaction between general constructions and specific instances of those constructions in the histories of the Spanish *estar* progressive and the English
auxiliary construction with *can*. New lexically-particular instances of the grammaticalizing *[tener que + Infinitive]* with high token frequencies emerge in the period from the 15th to the 18th century; these are discussed further in Sections 3.2.3 and 3.2.4.

**2.4.1. Evidence of alternation with *haber***

Given that the infinitives *dar* ‘to give’ and *comer* ‘to eat’ combine to account for 85% of the combined tokens of the *[tener + (X) + que + Infinitive]* construction, it is tempting to presume that these two verbs represent the inception of the *[tener que + Infinitive]* construction. That is, one might hypothesize that the path of grammaticalization for the target construction can be traced to nothing more than the confluence of *tener*, the relative pronoun *que*, and *comer* ‘to eat’ or *dar* ‘to give’.

However, the situation is not quite so clear for two reasons. The first complicating issue is that the *[tener + (X) + que + Infinitive]* is yet another context in which *tener* alternates with *haber* during this time period. Tokens of the construction with *haber* ‘to have’ are less frequent in the corpus than those with *tener*, but they appear nonetheless:

(68)  *et ellos non an que comer nin traen armas*
And they NEG have.PRS.3PL REL eat.INF nor bring.PRS.3PL weapons
‘and they do not have anything to eat nor do they carry any weapons’
[13th c.; *GEI*: II, 134]

(69)  *Si tu ouisses queles dar ellos te podrjan leuar*
‘If you had (something) to give them, they could take you’
[13th c.; Davies 2002-]
Given this additional context of alternation with *haber*, it is possible that the construction with *tener* emerged from an identical one with *haber*. However, a cursory search of Davies’ *Corpus de Español*, suggests that tokens like those in (68-69) are actually less frequent than those with *tener*. This is apparently contrary to Yllera’s assertion that examples with *tener* appear later than those with *haber* (1980:117). If *tener* had to first supplant *haber* in this context on its way to becoming more frequent than the latter, it is surprising that it would have done so already in such an early stage in the history of Spanish, especially given that *haber* continues to be the more frequent possession verb during the period analyzed here.

A thorough examination of *haber* itself is beyond the scope of the present investigation. Either explanation could only be confirmed by an analysis of texts even older than those utilized here. Earlier documents are not readily available, nor would they necessarily be classifiable as Spanish, as the socio-political events and linguistic standardization that brought about the variety known as “Spanish” (Castilian) had not progressed sufficiently by the 12th century to do so.

### 2.4.2. Alternative constructions with *tener*

It is, in fact, likely that the situation is even more complicated than a simple replacement of *haber* by *tener*. In addition to alternating with *haber* in the [*haber/tener* + (X) + *que* + Infinitive] construction, *tener* appears in several other, similar constructions in the corpus. These are structurally very similar, and vary only in the presence or absence of the relative pronoun *que* and/or a preposition. That is to say, this group of constructions might be described as fitting into a single schematic [*tener* + (Obj) + (Prep) + (*que*) + Infinitive].

---

16 For a complete list of all of the related constructions in which *haber* appeared, see Yllera (1980: pp. 92-110).
This schematic suggests something noteworthy about the status of this initial stage of [‘tener que’ + Infinitive]: it is seemingly quite variable\(^{17}\). The examples below illustrate this variability:

\[tener\ en\ que\ +\ Inf.\]

(70) e el cauallo ... moriosele a cabo de dies dias,

and ART horse die.PST.3SG-REFL-DAT at end of ten days

e non tenia cauallo en que yr.

and NEG have.IPFV.3SG horse on REL go.INF
‘and the horse died after ten days, and he did not have a horse on which to go’

\[14^{\text{th}}\ c.;\ Zifar\]

\[tener\ de\ que\ +\ Inf.\]

(71) leuauan los barbaros su fijo catiuo et no
take_away.IPFV.3PL ART barbarians POSS son captive and NEG

\text{tenie} \quad de\ que\ quitar\ lo

have.IPFV.3SG of REL remove.INF OBJ
‘the barbarians took her son captive and she did not have (anything) with which to ransom him’ \[13^{\text{th}}\ c.;\ EE1: \text{folio 125v}\]

\[tener\ de\ +\ Inf.\]

(72) don Yllán dixo a-l Papa que pues al non

sir Yllán say.PST.3SG to-the Pope COMP then other NEG

\(^{17}\) Yllera (1980) also lists [‘tener a’ + Infinitive] as an attested construction, but no tokens were found here.
tenía de comer

have.IPFV.3SG of eat.INF

‘Sir Yllán said to the Pope that then he did not have anything else to eat’

[14th c.; Lucanor: 98]

Although they are superficially similar, a close examination of these examples reveals some important differences. In (70), the infinitive yr ‘to go’ is not a transitive verb. Therefore, its action is not performed on the possessed object of tener, which is cauallo ‘horse’. If we assume that as the [tener que + Infinitive] construction grammaticalizes, the object goes from being interpreted as the object of tener to being the object of the infinitive, this is not a context which would allow for the change. Indeed, all of the 14 tokens of the target construction found here include a transitive infinitive18. The example in (71) does exhibit a transitive infinitive, but again the action of that verb quitar ‘to get (him) out’ is not performed on the object of tener. Instead, the object of quitar is the widow’s son, rather than the implied money required to ransom him.

The token in (72) is the only example of the three above that is semantically identical to the target construction. The relative pronoun que is absent and the preposition de, which typically means ‘of’ or ‘from’, appears in its place. Also, comar ‘to eat’ commonly appears as the infinitive in the [tener que + Infinitive] tokens, so there is no reason to believe that the intended meaning here is anything other than would be conveyed by the target construction.

However, this example is the only appearance (in the present corpus) of the construction with the preposition de in place of the relative pronoun que. With so few tokens in these centuries, it is difficult to speculate here about the potential role of uses like those in (70-72) as the [tener que + Infinitive] construction grammaticalizes. The target construction itself occurs only 14 times in the documents analyzed here, and these alternatives are even less frequent.

18 The same can be said for the 40 supplemental tokens drawn from Davies’ (2002-) corpus.
2.4.3. Semantic considerations

For these earliest appearances of the \([tener \ que + \text{Infinitive}]\) construction in the 12\textsuperscript{th}, 13\textsuperscript{th} and 14\textsuperscript{th} centuries, the meaning of tener is always one of possession (or at least an ambiguous use of its earlier ‘hold’ meaning). This is not to say, however, that there are not subtle differences in the semantic interpretations of the individual tokens. Compare the two uses of the construction in (73-74):

(73)  \textit{Yo só omne estraño en esta çibdat , et tengo otros tres conpañeros, et non tenemos ninguna cosa que comer.}  

‘I am a stranger in this city, and I have three other companions, and we do not \textbf{have anything to eat.}’ [13\textsuperscript{th} c.; \textit{Calila}: 326]

(74)  \textit{et los que no tenien oro o plata que dar alos cogedores daquel pecho; dauan de qual quier de los otros metales}  

‘…and those who did not \textbf{have gold or silver to give} to the collectors of that tribute gave some other metal’ [13\textsuperscript{th} c.; \textit{EEI}: folio 62r]

In both cases, tener indicates (negated) possession of an object, and that object is the entity upon which the action of the infinitive is to be performed. But while there is no overt obligation in either instance, there is physiological necessity in (73) and implicit external obligation in (74). It is then suggestive that the two infinitives, \textit{comer} ‘to eat’ and \textit{dar} ‘to give’, account for 11 of the 14 total tokens of the target construction. To the extent that we can assume some amount of implied obligation associated with the contexts in which these infinitives occur, there may be an inferred link between the construction itself and that notion of responsibility or obligation. Once this sense of responsibility is conventionalized and assumed to be an inherent
component of the construction independent of the specific context, new sets of lexical infinitives may appear in the construction (see Section 3.2).

The tokens in (73-74) also illustrate an additional element that may play a role in the semantic development of tener in the [tener que + Infinitive] construction. Both examples appear in negated clauses, and it is possible that when the possession encoded by tener is negated, the bleaching, or at least backgrounding, of that lexical meaning is facilitated. Negation, potentially combined with the semantic contribution of verbs like comer ‘to eat’ and dar ‘to give’, would then create a favorable environment for the grammaticalization of the construction. Interestingly, of the total combined tokens of [tener que + Infinitive] that were extracted for this time period, 80% (43/54) were negated, as in (73-74) above. The extremely high rate of tokens with negative polarity suggests that in addition the potential semantic role played by negation, the very tener que construction itself may have originated in predominantly negative contexts.

2.5. Conclusion

By analyzing the numerous distinct meanings and contexts of use for the verb tener ‘to hold’, ‘to have’, it is possible to characterize its status both within and without the [tener que + Infinitive] construction in the early stages of the Spanish language. It is immediately apparent that tener is utilized to encode both the older sense of ‘hold’ and the innovative ‘possess’ meaning. Although it is evidently used to express possession, or at least ambiguous senses of ‘hold’, tener is clearly not yet established as the dominant form for general possession. The earlier form, haber, persists in this semantic domain and alternates with tener in several specific contexts, including the structurally congruent [haber + (Obj) + que + Infinitive].

Furthermore, abstract uses of tener are common, including tokens meaning ‘to believe’ and ‘to esteem’. The former type presents important evidence that it is not simply the adjacency
of tener and que that allows for the grammaticalization of tener. Tener que ‘to believe that’ is much more frequent than the construction which ultimately grammaticalizes as a periphrastic modal expression. Thus, although the lexical elements tener and que co-occur quite frequently in the tener que ‘to believe that’ usage, the mere adjacency of these forms is not sufficient to produce the requisite context for the grammaticalization of tener into a modal verb.

Bybee and Torres Cacoullos (2009) argue that grammaticalization is not uniform for all instances or subclasses of a given construction, and there is evidence to support that argument here. The specific instances (prefabs) of the construction with comer ‘to eat’ and dar ‘to give’ are by far the most frequent in the present corpus and in supplemental data examined here. They propel the grammaticalization of the general construction by facilitating its semantic change (the addition of a sense of responsibility) and allowing it to expand into a wider range of lexical infinitives.

Analyses of later time periods to be conducted hereafter will be guided by the evidence presented here. Specifically, the infinitive verb type with which the construction occurs will be scrutinized, given the preponderance of comer and dar in the 12th-14th centuries. The object of tener will also be of interest as a potential source of evidence that tener is being applied to more general notions of possession. Abstract objects, rather than the concrete examples commonly found here (e.g. food or money) would suggest that tener is losing its etymological source meaning ‘to hold’. The following chapters will determine and describe the quantitative and qualitative changes in the occurrence of the [tener que + Infinitive] construction as it increases in frequency and gradually shifts from indicating possession to expressing obligation.
Chapter 3

Semantic and structural changes in [tener que + Infinitive]:
the 15th-18th centuries

Following the 14th century, the development of the [tener que + Infinitive] construction is marked by a number of changes relating to its meaning and its form. Chief among these are a set of advancements in its frequency, evidently linked to the semantic and structural modifications that emerged as the construction has continued to grammaticalize. The intervening time between the earliest occurrences of [tener que + Infinitive] (see Chapter 2) and the point at which it is firmly established in its modern patterns of use is crucial to the evolution of the construction. In the 18th century, the [tener que + Infinitive] construction has acquired modal meaning, but has not yet reached a token frequency high enough to contest with the older modal obligation constructions ([haber de + Infinitive] and [deber (de) + Infinitive]). It is not until the 19th century that its frequency rivals the frequencies of the two older constructions, so the period from the 15th century to the 18th century is pivotal in the development of [tener que + Infinitive] (see Sections 1.4.2 and 1.4.3). A thorough examination of these centuries elucidates the path by which [tener que + Infinitive] evolved from expressing possession to expressing obligation.

Taking the status of the construction in Medieval Spanish as its point of departure, the present chapter treats each of the noteworthy changes in its meaning, form, and frequency in turn. The individual components of the manifold grammaticalization process(es) at work here are separated for the sake of explanation, but as will be discussed below, they are often overlapping and interdependent. The analysis begins with a discussion of the steady increase in the overall frequency of the construction as it becomes more and more common in written Spanish throughout this intermediate period. Differences among the various meanings for which [tener que + Infinitive] is used are then considered, as is the relationship between ongoing semantic
changes and the frequency with which the construction includes particular infinitive verbs. This is followed by an assessment of changes in the most common form taken by the construction, as several trends emerge in the placement (or absence) of a syntactic object of the infinitive verb in the construction. Additionally, the status of the relative pronoun *que* ‘which’ is considered, as its function in the construction also appears to change during this period.

### 3.1. Changes in the frequency of [*tener que* + Infinitive]

In Chapter 2, it was noted that examples of [*tener que* + Infinitive] occurred infrequently in the 12th, 13th, and 14th centuries. Such tokens were so infrequent, in fact, that it is difficult to make generalizations regarding the construction’s progress along a path of grammaticalization, aside from concluding that [*tener que* + Infinitive] includes only the conservative lexical meaning of *tener* in the construction’s infancy. It has been noted in the literature on grammaticalization that frequency plays an important role in the gradual acquisition of grammatical meaning by erstwhile lexical items. The spread of these items to previously unavailable contexts in the grammatical system is necessarily accompanied by an increase in their frequency of occurrence. (Bybee 2010:123). In the case of [*tener que* + Infinitive], this makes new contexts available for its appearance. Tokens in which *tener* continues to mean ‘to have’ persist in this time period (as in (75)), but infinitives that make the lexical meaning of *tener* unavailable begin to appear (76).

(75)  
\[ \text{no tienes} \ \hat{a}l \ que \ hacer \ ni \ pensar \]  
\text{NEG have.PRS.2SG other REL do.INF nor think.INF}  
‘you do not have anything else to do nor think about’  
[15th c.; *Corbacho*: pt. IV, ch. 2]
The example in (75) illustrates the earliest uses of the construction. The action of the transitive infinitive hacer ‘to do’ would be performed on the possessed object of tener ‘to have’, which in this case is the abstract ál ‘(something) other’. This is the common arrangement for the construction in the 12th-14th centuries. The possessed object (whether implied or explicit) is already in existence, and often concrete.

This contrasts with the 18th century example in (76), where the infinitive venir ‘to come’ is intransitive, meaning that there is no element that could be construed as the possessed object of tener (una misa ‘a mass’ is the object of cantar ‘to say’). Such a token of the construction is therefore incompatible with an interpretation in which tener denotes possession and much more easily yields an obligation reading. Evidently, [tener que + Infinitive] has been reinterpreted by speakers as a construction that need not necessarily be limited to encoding possession.

The appearance of [tener que + Infinitive] with intransitive infinitives entails an expansion of the range of contexts that are available to the construction, meaning that it can occur more frequently. This creates a sort of feedback effect: increased frequency contributes to a change in the construction’s semantic and functional status, which itself allows the construction to occur more frequently. This point will be addressed in more detail in the following sections.
3.1.1. Corpus and data extraction

Tokens from the 12th-14th centuries are extremely rare, but it is possible to confidently establish the origins of the construction in that period, as does Yllera (1980)19. Furthermore, this initial period in the genesis of [tener que + Infinitive] serves as a useful baseline for the semantic and structural arrangement from which the modern modal construction coalesces. With only 14 occurrences in the corpus compiled here for the 12th, 13th, and 14th centuries and 40 in Davies’ (2002-) online Corpus del Español, the target construction appears just 54 times in a combined total of 10,897,674 words (just under 5 tokens per 1,000,000 words in the combined corpus). To compare the overall frequency of [tener que + Infinitive] for this early period with that of the intermediate period analyzed in the present chapter, tokens of the target construction were again extracted exhaustively from a corpus of written documents from the 15th-18th centuries, presented in Table 3-1 below. References for the corpus appear in the appendix.

19 It is worth noting here that although occurrences of the target construction evidently do not appear earlier than the 12th century, linguistic innovations tend to appear in the spoken language earlier than in writing. However, the limitations of the linguistic evidence available from this time period mean that we have no choice but to base conclusions on written texts.
Table 3.1. Corpus of 15th-18th century texts and [tener que + Infinitive] tokens extracted from each.

<table>
<thead>
<tr>
<th>Century</th>
<th>Words</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>15th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>La Celestina</td>
<td>85,180</td>
<td>11</td>
</tr>
<tr>
<td>El Corbacho</td>
<td>75,364</td>
<td>5</td>
</tr>
<tr>
<td>Century total</td>
<td>160,544</td>
<td>16</td>
</tr>
<tr>
<td>16th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lazarillo de Tormes</td>
<td>21,817</td>
<td>4</td>
</tr>
<tr>
<td>Crotalón</td>
<td>109,492</td>
<td>10</td>
</tr>
<tr>
<td>Segunda Celestina</td>
<td>101,261</td>
<td>10</td>
</tr>
<tr>
<td>DLNE (16th century)</td>
<td>85,939</td>
<td>17</td>
</tr>
<tr>
<td>Century total</td>
<td>318,509</td>
<td>41</td>
</tr>
<tr>
<td>17th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quijote (segunda parte)</td>
<td>167,174</td>
<td>18</td>
</tr>
<tr>
<td>DLNE (17th century)</td>
<td>88,571</td>
<td>12</td>
</tr>
<tr>
<td>Century total</td>
<td>255,745</td>
<td>30</td>
</tr>
<tr>
<td>18th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comedia Nueva</td>
<td>14,039</td>
<td>5</td>
</tr>
<tr>
<td>El sí de las niñas</td>
<td>20,019</td>
<td>7</td>
</tr>
<tr>
<td>DLNE (18th century)</td>
<td>74,901</td>
<td>19</td>
</tr>
<tr>
<td>Century total</td>
<td>108,959</td>
<td>31</td>
</tr>
<tr>
<td>Grand total</td>
<td>843,757</td>
<td>118</td>
</tr>
</tbody>
</table>

3.1.2 Frequency of occurrence for [tener que + Infinitive] in the 15th-18th centuries

In looking at the numbers of tokens of [tener que + Infinitive] found for these centuries, two important observations are immediately apparent. First, the overall frequency of the construction has increased dramatically from the extremely low average rate at which it occurred in the 12th-14th centuries. The second observation is that despite this increase, the frequency of occurrence remains relatively low, providing only 118 tokens for the four centuries in the period discussed here. Figure 3-1 illustrates this, comparing the average rate for the 12th-14th century data with the frequency of occurrence found in the corpus for each of the 15th, 16th, 17th, and 18th centuries.
The average frequency for the 12\textsuperscript{th}-14\textsuperscript{th} centuries is only 5 tokens per 1,000,000 words, but for the 15\textsuperscript{th} century, it climbs to 100. The rate remains almost constant for the 16\textsuperscript{th} and 17\textsuperscript{th} centuries, for which it is 129 and 117, respectively. Then in the 18\textsuperscript{th} century, the construction appears in the corpus at a frequency of 284 tokens per million words, a major increase compared to the almost nonexistent rate for the first three centuries. Despite the relative spike in frequency for \textit{tener que} + Infinitive for the four centuries examined here, the construction’s rate of occurrence never climbs above 300 tokens per 1,000,000 words. This is extremely low when compared with the overall rate for \textit{tener} itself, which appears 2586 times in the corpus used to analyze the 12\textsuperscript{th}, 13\textsuperscript{th} and 14\textsuperscript{th} centuries in Chapter 2, totaling 1,150,700 words. Over those three centuries, the rate of appearance for \textit{tener} is 2247 tokens per 1,000,000 words, meaning that \textit{tener} occurs nearly 450 times for every token of the target construction in that time period.
A limitation of the corpus compiled for the intermediate period examined in the present chapter is that the number of words for each individual century never exceeds a few hundred thousand, meaning that the upward trend in frequency suggested by Figure 3-1 is susceptible to being influenced by single documents for which the rate is unusually high (e.g. the frequency for the 16th century portion of the DLNE is 198 tokens/1,000,000 words, considerably higher than the overall frequency of 124 for that century). To combat this and more firmly establish the apparent trend of increasing frequency for the target construction, Davies’ (2002-) Corpus del Español was employed once again as a secondary source of data. Tokens of [tener que + Infinitive] were extracted exhaustively from a selection of texts from the corpus for the 15th, 16th, 17th, and 18th centuries. Table 3-2 displays the size of the supplemental corpus and the token counts for the data extracted from it.

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># of documents</td>
<td>42</td>
<td>61</td>
<td>70</td>
<td>41</td>
<td>214</td>
</tr>
<tr>
<td># of words</td>
<td>8,090,490</td>
<td>4,180,411</td>
<td>2,895,919</td>
<td>3,197,275</td>
<td>18,364,095</td>
</tr>
<tr>
<td>Tokens extracted</td>
<td>65</td>
<td>283</td>
<td>299</td>
<td>378</td>
<td>1025</td>
</tr>
</tbody>
</table>

The advantage of supplementing the present analysis with data from the large online corpus is that it allows for ameliorating the potential skewing effects of individual documents for which the rate of occurrence of [tener que + Infinitive] is unusually high. The word counts for the supplemental corpus in each century are high, as are the numbers of documents from which the data are taken. This also serves to minimize any effects of specific genres or authors that may employ the target construction at higher than average rates.

Figure 3-2 plots the frequency of occurrence of [tener que + Infinitive] for each century, this time for the supplemental data taken from Davies’ corpus.
The graph in Figure 3-1 is distinct from that in Figure 3-2 in that the former depicts a sharp increase in the rate of occurrence from the 14th to the 15th century, whereas the latter exhibits a large leap in frequency between the 15th and 16th. Despite this difference, and the fact that the maximum rate for the supplemental corpus (119 tokens/1,000,000 words) never reaches that of the corpus compiled for the present study (284 tokens/1,000,000 words), it is without doubt that [tener que + Infinitive] undergoes a sizeable and consistent increase in frequency during this time period. In the four centuries following the 14th, it goes from being nearly nonexistent in the written language to being a conventionalized, repeated construction in the 18th. After combining the data illustrated Figures 3-1 and 3-2, this increase can be quantified as a 2480% expansion from the average rate of occurrence for [tener que + Infinitive] in the 12th-14th centuries (from 5 tokens to 124 tokens per million words). It is important to note that this increase in frequency is not simply due to a proportional increase in the overall frequency of the verb tener.
‘to have’ over these four centuries. The rates at which tener occurs in Davies’ corpus for the 13th to 18th century are plotted in Figure 3-3.

Figure 3-3. Overall frequency of occurrence of tener ‘to have’ in Davies’ (2002-) corpus from the 13th to the 18th century (number of tokens per million words)

Of note in Figure 3-3 is the sharp increase in frequency from the 15th century to the 16th century, strikingly similar to that for [tener que + Infinitive] over the same span. Presumably, as tener acquired more and more of the semantic territory once occupied by haber ‘to have’ (see Chapter 2 and Section 3.2 below), tener became more frequent, as did the particular construction analyzed here. The two plots differ, however, in the relative amplitude of that increase: where [tener que + Infinitive] is almost 25 times more frequent in the 16th century than in the 15th, the rate for tener in the 16th is only 2.2 times greater than that for the 15th. What is more, while [tener que + Infinitive] continues to increase in frequency over the final two centuries of this time period, tener actually becomes less frequent in Davies’ corpus. Tener itself has not become more
frequent in the latter half of the intermediate period discussed here, but the [tener que + Infinitive] construction evidently makes up an increasingly larger portion of the total occurrences of tener in those final two centuries.

3.2. Semantic changes evident in [tener que + Infinitive]

Along with an increase in frequency, another crucial component in the grammaticalization of [tener que + Infinitive] is the semantic development undergone by both the construction as a whole and the verb tener ‘to have’ within it. As discussed in Chapter 2, tener has not yet established itself as the dominant possession verb in Spanish by the end of the 14th century. In the first three centuries of the present analysis, its competitor haber ‘to have’ continues to occupy a number of semantic domains and lexical functions that will eventually become part of the jurisdiction of tener. In the 12th-14th centuries, haber was the more common of the two verbs for encoding basic possession, and occupied that role for expressing abstract possession of emotions and sensations (see Section 2.2.2). This holds true in the 15th century, as evidenced by (77).

(77)  

and of God NEG have.PRS.3SG fear nor dread

‘and of God (he) has neither fear nor dread’ [15th c.; Corbacho: pt. IV, ch. 2]

In the 16th century, the preponderance of haber appears to be waning, as tokens of miedo ‘fear’ occur with both haber and tener, even within the same document ((78) and (79), respectively).
(78)  *Venida la noche, y el no, yo hube miedo de quedar en casa solo*

‘Once arrived the night, and he not, I had fear of staying at home alone’

[16th c.; Lazarillo: 59]

(79)  “*Plega a Dios que no me muerda -decia yo-, que harto miedo le tengo.*”

“Pray to God that it does not bite me,” I was saying, “as I have much fear of it”

[16th c.; Lazarillo: 33]

This pattern of alternation between *haber* and *tener* continues through the 18th century in the documents included in the present corpus. But while tokens of *haber* with *miedo* ‘fear’ continue to occur in the 18th century, they are less frequent than tokens with *tener*. These occurrences cannot necessarily be taken as evidence that *tener* has displaced *haber* wholesale in the domain of possession, but it is certainly suggestive that *tener* has made inroads in expressions of emotion and sensation. Considering the sharp rise in frequency for [*tener que + Infinitive*] from the 15th to the 16th century (see Section 3.1) it stands to reason that an increase in frequency for *tener* in general would be accompanied by subsequent increases in frequency for individual constructions in which *tener* appears.

Importantly for the investigation of [*tener que + Infinitive*], examples like that in (5) illustrate the extension of the meaning of *tener* to the possession of abstract objects. Whereas *tener* originates from an etymological source limited to denoting the physical grasping of a concrete object, it is evident that by the 18th century, the verb has developed a wide range of functions for the expression of possession. This is a crucial element in the increase in frequency experienced by the [*tener que + Infinitive*] construction through the intermediate time period: as *tener* begins to be applied to abstract possession, it becomes possible for the target construction to include infinitive verbs that would otherwise be incompatible with the concrete etymological meaning of *tener*. 
3.2.1. Infinitive verbs frequently occurring within [tener que + Infinitive]

Evidence of the continuing abstraction and generalization of the meaning of tener is found in the types of infinitives with which [tener que + Infinitive] appears in the 15th-18th centuries, as well as the frequencies of these infinitives. To investigate the potential impact of commonly occurring infinitives on the grammaticalization of [tener que + Infinitive] and the role they play in the semantic evolution of the construction, all tokens of [tener que + Infinitive] found in the 15th-18th century corpus were coded for their included infinitive. Table 3-3 displays the distribution of the five most frequently occurring infinitives in the data extracted from the corpus for the 15th-18th centuries.\(^{20}\) When compared with the apparent tendencies of certain infinitives to appear in the data from the 12th-14th centuries, some noteworthy trends appear in this intermediate time period.

<table>
<thead>
<tr>
<th>Infinitive</th>
<th>Tokens per century</th>
<th>N/total N for time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>comer ‘to eat’</td>
<td>15(^{th}) 1 3 0 1</td>
<td>5/118 (4%)</td>
</tr>
<tr>
<td>dar ‘to give’</td>
<td>16(^{th}) 3 4 3 0</td>
<td>10/118 (8%)</td>
</tr>
<tr>
<td>decir ‘to say, to tell’</td>
<td>17(^{th}) 0 4 6 1</td>
<td>11/118 (9%)</td>
</tr>
<tr>
<td>hacer ‘to do’</td>
<td>18(^{th}) 4 7 5 5</td>
<td>21/118 (18%)</td>
</tr>
<tr>
<td>ver ‘to see’(^{21})</td>
<td>N/total N for century 8/16 (50%) 20/41 (49%) 20/30 (67%) 9/31 (29%) 57/118 (48%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-3. Distribution of the most common infinitive types for the 15th-18th centuries

\(^{20}\) The infinitive añadir ‘to add’ appeared in nine tokens, but was excluded from the list in Table 3-3 because its frequency is an effect of genre. All nine tokens of añadir are limited to the 17th and 18th centuries and eight of the nine are drawn from just two documents (the Documentos Linguísticos de la Nueva España for the 17th and 18th centuries). In these documents, añadir typically occurs in a list, as a specific recurring usage of the [tener que + Infinitive] construction used to explain that a legal document has been reviewed and approved by a government official (e.g. dijo que estaba bien escrito...y que no tenía que añadir, quitar ni enmendar ‘[he] said that it was well written and that he had nothing to add, remove, nor amend’ [[17th c.; DLNE: 402].

\(^{21}\) The lexical meaning ‘to see’ is supplied here for ver, but as will be discussed in Section 3.2.4, ver used with [tener que + Infinitive] takes on a meaning similar to that of hacer, hence the common usage of tener que ver (con) ‘to have to do (with)’.
It is immediately evident in Table 3-3 that these five verbs consistently account for a substantial portion of the data, combining to make up 48% (57/118) of the data in the corpus for this time period, and remaining near 50% for the first three of the four centuries. But upon closer inspection, the predominance of the two most frequent infinitives from the 12th-14th century data has shifted.

In Chapter 2, it was noted that dar ‘to give’ and comer ‘to eat’ accounted for 44% and 41% of the data, respectively. This contrasts with the data from the intermediate time period, where comer only occurs five times. These five tokens are enough to make comer one of the five most frequent infinitives from the 15th to the 18th century, but it has clearly been surpassed by the other four verbs considered here. This is especially true for hacer ‘to do’, which is the single most frequent infinitive in the data. The prevalence of hacer evidently has implications for the semantic evolution of the target construction; this is discussed below in Section 3.2.3. While the total number of tokens is limited by the size of the corpus and the fact that [tener que + Infinitive] is relatively infrequent, the pattern of high frequencies for the five infinitives in Table 3-3 is borne out in the supplemental data extracted from Davies’ Corpus del Español (2002-).

To more firmly establish this pattern, the additional tokens from Davies’ corpus were also coded for the infinitive included in the [tener que + Infinitive] construction. When the five verbs displayed in Table 3-3 are examined in the supplemental data extracted from Davies’ corpus, it is possible to make additional generalizations about their occurrence in the target construction, as the pattern is nearly identical to that found in the corpus compiled here. Table 3-4 demonstrates this concurrence.
Table 3-4. Distribution of common infinitive types for the 15\textsuperscript{th}-18\textsuperscript{th} centuries in Davies’ (2002-) corpus

<table>
<thead>
<tr>
<th>Infinitive</th>
<th>15\textsuperscript{th}</th>
<th>16\textsuperscript{th}</th>
<th>17\textsuperscript{th}</th>
<th>18\textsuperscript{th}</th>
<th>N/total N for time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>comer ‘to eat’</td>
<td>17</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>32/1025 (3%)</td>
</tr>
<tr>
<td>dar ‘to give’</td>
<td>7</td>
<td>20</td>
<td>23</td>
<td>15</td>
<td>65/1025 (6%)</td>
</tr>
<tr>
<td>decir ‘to say, to tell’</td>
<td>4</td>
<td>29</td>
<td>17</td>
<td>19</td>
<td>69/1025 (7%)</td>
</tr>
<tr>
<td>hacer ‘to do’</td>
<td>14</td>
<td>46</td>
<td>41</td>
<td>38</td>
<td>139/1025 (14%)</td>
</tr>
<tr>
<td>ver ‘to see’</td>
<td>2</td>
<td>24</td>
<td>19</td>
<td>25</td>
<td>70/1025 (7%)</td>
</tr>
<tr>
<td>N/total N for century</td>
<td>44/65 (68%)</td>
<td>130/283 (46%)</td>
<td>102/299 (34%)</td>
<td>99/378 (26%)</td>
<td>375/1025 (37%)</td>
</tr>
</tbody>
</table>

These supplemental data confirm the generalizations drawn from the data presented above in Table 3-3, as well as revealing some intriguing patterns in the distribution of these frequent infinitives. The assertion that these five infinitives account for a sizable portion of the data is cemented, as 37\% of the data (375 of 1025 total tokens) from this time period include one of the five. What is more, they illustrate that the pattern in the original data is effectively identical to that found here, including the fact that the occurrence of comer ‘to eat’ decreases greatly in frequency after the 16\textsuperscript{th} century, dropping to only two tokens in each of the 17\textsuperscript{th} and 18\textsuperscript{th} centuries. Although this infinitive provided a sort of starting point for the construction in the 12\textsuperscript{th}-14\textsuperscript{th} centuries, it appears that [tener que + Infinitive] has achieved autonomy as a construction and no longer relies on comer as a specifically frequent form to support it.

The waning of [tener que + comer] ‘to have (food) to eat’ as a commonly recurring specific instance may actually contribute to the semantic evolution of the general construction. Comer has been surpassed in frequency by several other infinitives in the 17\textsuperscript{th} century and has dropped off of the list of the five most frequent infinitives appearing in the construction. The four remaining verbs on that list (from Tables 3-3 and 3-4) share a common thread: dar ‘to give’, decir ‘to say’, hacer ‘to do’, and ver ‘to see’ all include a sense of intention, motivation, and/or
externally imposed responsibility when used in the target construction, as demonstrated in examples (80-83).

dar ‘to give’:

(80) *Yo no tengo dineros que os dar por la llau*

‘I do not have money to give you for the key’ [16th c.; Lazarillo: 26]

decir ‘to say’

(81) *Y respondio avia ya dicho todos sus peccados y confesadosse enteramente, y que no tenia más que decir*

‘And (he) responded (that) he had already told all of his sins and confessed himself completely, and that he had no more to say’ [17th c.; DLNE: 261]

hacer ‘to do’

(82) *En mi teniades bien que hacer, y no hariades poco si me remediaseses*

‘In me you would have much to do, and no little thing would you do if you remedied me’ [16th c.; Lazarillo: 25]

ver ‘to see’

(83) *¿Quiéreslo ver? Pues si miras en esto, no tenia más que ver la letra désta con la del otro día que yo con el rey.*

‘Do you want to see it? Well, if you look at it, the handwriting of this one does not have any more to do with that of the other than I do with the king.’ [16th c.; Segunda Celestina]

In each case, the infinitive imparts a loosely modal element to the meaning of the construction. For example, in (80) and other tokens with *dar* ‘to give’, the context commonly indicates a motivation for the action of *dar*, as with a debt, an exchange of goods, or a donation to a person who is in need. Similarly, the sinner in (81) is left with nothing to confess, that is, he has nothing more he wishes or needs to say after making his confession. Of these four verbs, *hacer* ‘to do’ is the infinitive most closely linked with a sense of obligation. Tokens like that in (82)
typically involve a task or chore for which the subject of tener is responsible, implying a compulsion to complete that action.

This is akin to the interpretation associated with ver in [tener que + ver]. While the canonical definition of ver is ‘to see’ the verb covers a range of senses including ‘to meet, to visit, to deal with, to do with’, the latter being the conventionalized meaning associated with [tener que ver]. The example in (83) portrays that sense by referring to one entity’s involvement or association with another entity or event, albeit with a less overt indication of responsibility or obligation. However, the ‘have to do’ meaning of tener que ver is not unlike that of tener que hacer, again implying a slight sense of commitment or engagement, as with metaphorically “seeing to” the completion of an errand. Due to their evident impact on the grammaticalization of [tener que + Infinitive], the infinitives ver and hacer are treated in more detail in Sections 3.2.3 and 3.2.4, respectively.

3.2.2. Bridging contexts

The semantic element of responsibility, intention, or compulsion that is common to tokens featuring the four verbs discussed above allows for a reinterpretation of the overall meaning of the construction. When reinterpretation of a form’s meaning is available in a particular context, the construction is presumed to occupy what has been described as a ‘bridging context’ (Evans and Wilkins 1998; Heine 2002:84). Based on this perspective, the usage of a form (e.g. the [tener que + Infinitive] construction) in a particular context may trigger an inference on the part of a listener that in addition to the source meaning, there is another plausible interpretation. This is not to say that the source meaning is no longer available, but in the right context, the innovative or “target” meaning may actually be more plausible. The example in (84) illustrates this availability of inference.
...cuando vuestra merced quisiere ahorrar caminos y trabajos para llegar a la inaccesible cumbre del templo de la Fama, no tiene que hacer otra cosa sino dejar a una parte la senda de la poesía, algo estrecha, y tomar la estrechísima de la andante caballería

...if you wish to spare yourself odysseys and labors in reaching the inaccessible summit of the temple of fame, you have to do no other thing but to leave aside the somewhat narrow path of poetry and take the even narrower one of knight-errantry [17th C.; Quijote: 780]

The source meaning of the [tener que + Infinitive] construction is clearly available here, in that the NP otra cosa ‘other thing’ may be interpreted as a possessed object of tener ‘to have’: no tiene que hacer otra cosa ‘you have to do no other thing’. However, this is not the only interpretation available. It is also possible to analyze otra cosa as the patient of hacer, as the action of the infinitive (‘to do’) would ultimately be performed on it. What is more, the fact that the clause is negated may suggest to a listener/reader that an inference of obligation is appropriate, as the ‘other thing’ is both abstract and its possession is semantically negated.

Perhaps most importantly, the larger context in which the construction appears is also a contributing factor to the potential for inference. The advice provided by the speaker regards a necessary means to accomplish a desired goal, precisely the sort of externally imposed obligation that can foreground the target (i.e. modal) meaning. Thus, the potential presents itself for the inference that the intended meaning is ‘you do not have to do any other thing’ instead of ‘you do not have any other thing to do’, despite the fact that both are viable. As Evans and Wilkins put it:

speech participants do not detect any problem of different assignments of meaning to the form because both speaker and addressee interpretations [sic] of the utterance in context are effectively, functionally equivalent (if semantically distinct). (1998:5)
In this context, it is not necessary to infer one meaning or the other. Here they are both available, and this is sufficient for laying the groundwork for the semantic change. This is precisely what can ultimately result in the conventionalization of the pragmatic inference that the intended meaning is the target (obligation) meaning, a process integral to grammaticalization (Hopper and Traugott 1993: 75-77).

A variable that may contribute to the pragmatic inference that the modal meaning is intended is the polarity of the clause in which the construction appears. As discussed in Chapter 2, the rate of negated tokens was quite high in the 12th-14th century data; 43/54 tokens were negated (see Section 2.4.3). That rate has decreased by the end of the time period considered here, as only 30% (122/409) of the 18th century tokens are negated. Still, that means nearly a third of the 18th century data appear in negated contexts, which may still be enough to facilitate the backgrounding/bleaching of the lexical possession meaning of tener. For example, with (84) above, no tiene que hacer otra cosa ‘you have to do no other thing’, could be interpreted as ‘you do not have to do any other thing’. This may not fundamentally change the meaning, but assuming that the entire construction is negated rather than just the object may foreground the obligation meaning of the tener que construction. Negation as a conditioning factor in the variation between [tener que + Infinitive] and other modal constructions is examined in Chapters 4 and 5.

3.2.3. tener que hacer

An additional variable that appears to be influential in the grammaticalization of [tener que + Infinitive] is the occurrence of highly frequent, lexically-specific instances of the construction. The most frequently occurring infinitive in both the original corpus and the supplemental data drawn from Davies’ corpus, hacer ‘to do’ appears in the [tener que +
Infinitive] construction twice as often as the second most frequent verb, *ver* ‘to see’. Accounting for 14% (160 of 1143 total tokens) of the data for this time period in the two corpora combined, the high frequency and semantic association of *hacer* with intention and responsibility evidently contribute to the evolving meaning of the target construction as a whole. That is, the meanings that may be inferred for many occurrences of [*tener que* + Infinitive] with *hacer* as the infinitive appear to be an avenue for the inference of obligation as the construction grammaticalizes.

The majority of tokens of [*tener que* + *hacer*] in this time period can be characterized as belonging to one of two basic categories. The first of these is the set of tokens for which there is an overtly expressed object, such that an interpretation of *tener* as indicating possession is justifiable on both syntactic and semantic grounds. Examples (85-86) depict this type of token.

(85)  *como si no tuviera otra cosa que hacer*

‘as if (he) did not have anything else to do’

[16th c.; Davies 2002-]

(86)  *tener negocios que hacer en Cartagena*

‘having business to do in Cartagena’

[17th c.; Davies 2002-]

These instances clearly illustrate the lexical (possession) meaning from which the construction originates, but they are less frequent than the type portrayed in (87-88). Here there is still an overtly expressed object, but it is less specific.

(87)  *En mi teniades bien que hacer, y hariades poco si me remediasedes*

‘In me you would have much to do, and no little thing would you do if you remedied me’ [16th c.; Lazarillo: 25]
(88) Harto tengo, hija, que llorar, acordándome de tan alegre tiempo y tal vida como yo tenía

‘Much I have, daughter, to lament, recalling the happy times and such life as I had’ [15th c.; Celestina: 236]

Tokens with this type of non-specific object make up the majority of examples with an overt object. In addition to bien ‘much’ and harto ‘much’, a number of other nouns expressing quantity appear in these tokens, including mucho ‘a lot’, más ‘more’, tanto ‘so much’, menos ‘less’, and poco ‘little’. In fact, these tokens lead to a somewhat ambiguous reading,[22] as teniades bien que hacer in (87) could arguably be interpreted as ‘you would have to do a lot’ or ‘you would have a lot to do’.

Slightly more ambiguous are the second basic type of hacer tokens: those for which there is no overtly expressed object. In these instances, the meaning is still evidently one of possession, but as with the other hacer tokens, there is undoubtedly an implication of purpose, intention, or responsibility.

(89) Verdad es, pero mal conosci a Celestina. Quando ella tiene que hacer, no se acuerda de Dios ni cura de santidades.

‘That is true, but you know Celestina poorly. When she has (something) to do, she worries neither about God nor sanctity.’ [15th c.; Celestina: 222]

(90) ...es muy doloroso asistir a tan acerbo espectáculo; tengo que hacer.

‘It is very upsetting to be here for such a harsh spectacle; I have (things) to do.’ [18th c.; CN: 122]

The examples in (89) and (90) demonstrate the common use of [tener que + hacer] during this time period to describe a commitment, errand, or chore for which the subject of tener is

[22] The ambiguity here is assumed to be in the syntactic-semantic sense suggested by Evans and Wilkins (1998), rather than functional ambiguity.
responsible. Here, the object is implicit and undefined (‘something’), allowing for the inference that the speaker intends to foreground the notion of obligation.

An additional class of [tener que + hacer] tokens emerges during this time period, first appearing in the 17th century. These are tokens in which there is sometimes an expressed object, but the implication of obligation appears to be even stronger than in those lacking an overt object.

(91)  *lo dijo al padre Villarás, y le advirtió que, si tenía que hacer testamento, lo hiciese con brevedad*

‘he told Father Villarás and advised him if he had to make a will, that he do it soon’
[17th c.; Davies 2002-]

(92)  *Mi padre me dirá en esto lo que tengo que hacer*

‘My Lord will tell me then what I have to do’
[17th c.; Davies 2002-]

In both (91) and (92), contextual circumstances suggest an interpretation of the construction as indicating obligation rather than possession of any kind. The priest in (17) is dying, and is being instructed by his doctor to make his will as soon as possible; that is, if the priest wishes to or needs to. The speaker in (92) is also anticipating instructions, this time from God, regarding what he is to do. In each case, the token lacks an object that could easily be construed as the possessed object of tener. The unspecified lo que ‘what’ in (18) refers to some action that has yet to be determined, and the will in (91) does not yet exist.

What is more, the word order in (91) is potentially used to signal the foregrounding of the obligation component, given that testamento ‘will’ is postposed, appearing after the infinitive

---

23 Tokens like those in (15) and (16) contrast slightly with tokens in which the relative pronoun que ‘which’ is represented orthographically as qué. This orthographic difference (the accented é) suggests that the pronoun would be stressed in speech, whereas in (15) and (16) it would not (e.g. *Del linaje plebeyo no tengo qué decir* ‘About the plebian lineage I do not have anything to say’). A lack of stress on the unaccented que could be taken as evidence of backgrounding the notion of possession, but as the written materials utilized for the present corpus do not allow for prosodic analysis, a more detailed discussion of this issue is omitted here.
(compare English *I have something to do* vs. *I have to do something*). This is in contrast with the majority of the tokens for which there is an overt object, where that object either appears before *tener*, as in (88) or intervenes between *tener* and *que* in the construction (85-87). Changes in the order of the elements in the [*tener que* + Infinitive] and their role in the grammaticalization of the construction will be discussed in more detail in Section 3.3.

Given that *hacer* is the most frequently occurring infinitive in the target construction for these four centuries, the semantic interpretations commonly associated with its use are likely to have a larger impact on the conventionalized meaning of the construction than other verb types. To illustrate the influence of frequency on conventionalized structures and meanings, one need not look beyond the common co-occurrence of the relative pronoun *que* and the infinitive *hacer*. As mentioned above, these two lexical elements appear alongside one another quite frequently in the data analyzed here, but their co-occurrence is not limited to [*tener que* + Infinitive]. In fact, *hacer* immediately follows *que* some 436 times in Davies’ (2002-) corpus for the 15th century alone; of these, 65 are tokens of [*tener que* + *hacer*]. Eventually, the pronoun + verb collocation was reanalyzed as a single word, *quehacer* ‘chore’ (literally ‘something to do’), with effectively the same meaning as that found in (89) and (90) above (Penny 2002: 301).24

### 3.2.4. *tener que ver*

The second most frequent infinitive in the data is perhaps not as influential in the semantic development of [*tener que* + Infinitive], but warrants discussion nonetheless25. *Ver* ‘to see’ appears in the target construction a total of 80 times in the data from the combined corpora.

24 Incidentally, the following token was found in the 20th century corpus (see Chapter 5), juxtaposing the fused *quehacer*, in this case transcribed as *quihacer*, with the *haber que* + Infinitive construction (Section 4.1.1): Porque, pus... ya ve usté que **hay que hacer el quehacer a la hora que se necesita**. ‘Because, well... you see that **you have to do the chore** at the time that it is needed.’ [MexPop]

25 The infinitive *decir* ‘to say’ appears exactly as many times in the data as *ver* ‘to see’. However, because the latter offers a more intriguing contribution to the analysis, it is treated in detail here.
Thus, it is exactly half as frequent as hacer ‘to do’. This specific usage of the target construction is effectively synonymous with English to have to do (with); that is, ‘to be associated with’.

Presumably, the link between ‘to see’ and the meaning associated with such instances of the construction is the extension of ‘to see’ to ‘to visit’, and from there ‘to have business with’ and the more general ‘to be associated with’. Not surprisingly, a few tokens of the target construction with hacer ‘to do’ appear to convey this same sense:

(93)  *Pues si miras en esto, no tenía más que ver la letra désta con la del otro día que yo con el rey.* [16th c.; Segunda Celestina]

‘Well if you look at it, the handwriting from this [letter] doesn’t have any more to do with the one from the other day than I do with the king.’

(94)  *Avísame si el mayordomo que está contigo tuvo que ver en las acciones de la Trifaldi, como tú sospechaste* [17th C.; Quijote: 1050]

‘Let me know if the caretaker that is with you had an involvement in the events of Trifaldi, as you suspected.’

Of interest here, however, is the link between tokens of [tener que + ver] and an assertion made by Heine regarding semantically ambiguous contexts. He states: “bridging contexts may, but need not, give rise to conventional grammatical meanings” (2002: 85). Indeed, the general [tener que + Infinitive] construction has evolved into a periphrastic modal construction in Modern Spanish, but the specific instance tener que ver has largely maintained the ‘have an association with’
meaning it already had in the time period under consideration here. The examples in (93) and (94) are drawn from 16\textsuperscript{th} and 17\textsuperscript{th} century texts respectively, while (96) below is from the 20\textsuperscript{th} century.

(96) \textit{La xenofobia parece que poco tiene que ver con esta mezcla de sangres y culturas.} [20\textsuperscript{th} c.; Davies 2002-]

‘Xenophobia seems like it \textbf{has little to do} with this mixing of bloodlines and cultures.’

The lexically-specific \textit{tener que + ver} instance of the general \textit{tener que} construction allows for further reduction or bleaching of the possession sense of \textit{tener}. The meaning of the prefab is interpreted holistically as ‘to be associated with’, which minimizes the independent meaning contribution (i.e. possession) made by \textit{tener}. As with the tokens including \textit{comer} ‘to eat’ and \textit{dar} ‘to give’ in the 12\textsuperscript{th}-14\textsuperscript{th} century period (discussed in Section 2.4) and the frequent \textit{tener que hacer} (Section 3.2.3 above) we see more evidence that a lexically-specific instance of the construction can pave the way for the association of a more general sense of responsibility with the general \textit{tener que} construction.

Furthermore, tokens of \textit{tener que + ver} provide evidence that the general construction did not grammaticalize wholesale, given that \textit{tener que ver} ‘to have to do with, to be associated with’ has become a kind of fossilized offshoot of \textit{tener que + Infinitive}, preserving the same meaning in the modern language that it had in the 16\textsuperscript{th} century. It appears as though \textit{tener que ver} served an important purpose at this stage in the development of \textit{tener que + Infinitive}, acting as a booster rocket of sorts and providing momentum for semantic change in the general construction and ultimately being left behind. This development from one semantic stage to another is considered in more detail in the following section.
3.2.5. Semantic stages evident in the 15th-18th centuries

Returning to the discussion of the analogous English *to have to* construction in Chapter 2, it is possible to reevaluate the stages in its development proposed by Brinton (1991) and Fischer (1994) and consider how they apply to the grammaticalization of [*tener que* + Infinitive]. Figure 3-4 presents these stages.

| Stage 1 | *have* in the construction *have* + OBJECT + INFINITIVE is a full lexical verb meaning ‘possess’, while the nominal object denotes what is possessed and the infinitive serves as an ‘adverbial adjunct of purpose’, as in *I have food to eat*. |
| Stage 2 | The meaning of *have* changes, with possession and obligation meanings existing simultaneously. Syntactically, the structure is unchanged, but the object is now viewed as a responsibility or duty to be accomplished by the subject of *have*, as in *I have work to do*. |
| Stage 3 | Semantically, *have* now expresses obligation exclusively, and the object and infinitive have been transposed (*have* + INFINITIVE + OBJECT). The nominal object is now the object of the infinitive rather than *have* as in *I have to do work*. |

Figure 3-4. Developmental stages in the evolution of the English *to have to* construction (drawn from Brinton 1991 and Fisher 1994)

The stages in Figure 3-4 are a useful means of categorizing and classifying the numerous different instantiations of the [*tener que* + Infinitive] construction found in the 15th-18th century data. Stage 1 corresponds to the earliest tokens of the construction in the 12th-14th centuries, while Stage 2 captures the ambiguous bridging contexts discussed in Sections 3.2.1 to 3.2.4. That is not to say, however, that the semantic evolution of the construction is best characterized as a sequence of discrete steps. In fact, diachronic change in grammaticalization processes is commonly said to be necessarily gradual, involving synchronic variation in form and gradience in meaning. (Hopper and Traugott 1993; Bybee 2010: 121).

This is certainly true for [*tener que* + Infinitive] in the intermediate time period considered here. Tokens corresponding to Stage 1 in Figure 3-4 are common in the 15th century and examples that appear to convey the modern modal meaning (Stage 3) of the construction
already appear in the 18th century. These meanings co-occur and appear alongside tokens for which the meaning is ambiguous as well. Simply put, the appearance of tokens of the more advanced stages (2 and 3) does not preclude the appearance of relatively less advanced stages (1 and 2). While examples in which [tener que + Infinitive] conveys possession do necessarily precede those for which it conveys obligation, the possession usage does not vanish with the advent of the modal meaning. The examples in (97-99), which demonstrate a range of different meanings, all appear in the 17th century:

(97)  **tener negocios que hacer en Cartagena**

‘**having business to do** in Cartagena; [17th c.; Davies 2002-]

(98)  **Harto tenía que hacer el socarrón de Sancho en disimular la risa, oyendo las sandeces de su amo**

‘The rogue Sancho **had to do much** to hide his laughter, hearing the foolishness of his master’ [17th c., Quijote: 710]

(99)  **no tenía que hacer la proposición** [...] ni aparecía proceso, testigos ni competencias algunas y sólo instaba la porfía de los acusadores insolentes

‘[he] did not **have to make a statement** [...] nor was there process, nor witnesses, nor arguments whatsoever, and he only encouraged the complaints of the insolent accusers’ [17th c.; Davies 2002-]

The use of [tener que + hacer] in (97) suggests an interpretation with at least an ambiguous sense of possession coupled with intention, as with several of the tokens with **hacer** ‘to do’ presented in 3.2.3. Having business, errands, or chores to do entails a degree of responsibility, but can be considered possession nonetheless. This is further indicated by the object **negocios** ‘business’ intervening between **tener** and **que**, providing a structural signal that the meaning intended for **tener** is more lexical than modal. The example in (98) also includes an overt object, although that object, **harto** ‘much, a lot’, precedes **tener**. Arguably, **harto tenía que**
hacer could be translated as either ‘he had much to do’ or ‘he had to do much’ in this case. But what distinguishes this token from the former is that context clearly drives an inference of obligation, as the subject’s need to appear respectful and the apparently intense humor of the situation required that he expend a great deal of effort to hide his laughter.

Contrasting with the previous two, the example in (99) allows for only an interpretation of obligation. The object la proposición ‘the statement’ is best considered the object of hacer ‘to do’ rather than tenía ‘he had’ (i.e. ‘he did not have to make a statement’ instead of ‘he did not have a statement to make’). The coexistence of examples such as these in such close temporal proximity is important evidence of two related phenomena that Hopper (1991) refers to as “Layering” and “Persistence”. He defines these terms as follows:

“Layering”:
Within a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist and interact with the new layers.

“Persistence”:
When a form undergoes grammaticization from a lexical to a grammatical function, so long as it is grammatically viable some traces of its original lexical meanings tend to adhere to it, and details of its lexical history may be reflected in constraints on its grammatical distribution. (1991: 22)

In this case, an older meaning of the [tener que + Infinitive] construction remains in use even as innovative senses appear. Thus, the appearance of ‘have to do X’ uses like (91) do not entail the disappearance of ‘have X to do’ uses like (97). This is evidence of Hopper’s “persistence”, while the emergence of clear obligation tokens of [tener que + Infinitive] as in (99) is linked with the principle of layering, as the tener que construction encroaches on the functional domain already occupied by older modal constructions in Spanish (see Sections 4.1-4.3).
3.2.6. Infinitive types as evidence of semantic change

In addition to the mere existence of innovative meanings alongside more conservative lexical meanings, semantic generalization of the construction is evident in the changing distribution of [tener que + Infinitive] during this time period. Specifically, the infinitive types appearing within the construction provide quantitative evidence pertinent to Hopper’s claim regarding constraints on the grammatical distribution of a grammaticalizing form. Revisiting the discussion of frequent infinitives above in Section 3.2.1, it is important to note that the four most frequently occurring infinitives in the data make up a smaller and smaller portion of the data for each century. As has already been mentioned, comer ‘to eat’ was the single most frequent infinitive in the data for the 15th century, but virtually disappeared by the 17th century. The disappearance of comer seems to be symptomatic of a general decline in relative frequency for the most frequent infinitives in the combined corpora; the four other most frequent infinitives decrease in relative frequency over this period. For example, hacer ‘to do’, which was the most frequent overall, makes up 22% of the data in the 15th century, 16% in the 16th, 14% in the 17th, and 11% in the 18th.

Although hacer ‘to do’, ver ‘to see’, decir ‘to say’, and dar ‘to give’ remain the four most frequent infinitives in the data even in the 18th century, it is clear that [tener que + Infinitive] has been generalized as a construction that may appear with a wider and wider range of infinitive types. That is, the data are more and more evenly distributed across a range of infinitives, most of which did not occur at all in the construction prior to the 15th century. The changing distribution is evident in the ratio of the number of distinct infinitive types to the total number of tokens for each century, depicted in Table 3-5. To highlight the change, the type/token ratio (the number of distinct infinitive types divided by the total number of tokens) for each century is presented
alongside the corresponding combined relative frequency for the four most frequent infinitives in Table 3-5.

Table 3-5. Type/token frequencies for infinitives in [tener que + Infinitive] and proportion of data composed of the four most frequent infinitives (dar, decir, hacer, ver) in the intermediate time period

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
</tr>
</thead>
<tbody>
<tr>
<td>% data comprising four most frequent infinitives</td>
<td>64% (52/81)</td>
<td>46% (150/324)</td>
<td>37% (122/329)</td>
<td>26% (108/409)</td>
</tr>
<tr>
<td>Type/token ratio (# distinct infinitive types/total N)</td>
<td>0.321 (26/81)</td>
<td>0.324 (105/324)</td>
<td>0.371 (122/329)</td>
<td>0.403 (165/409)</td>
</tr>
</tbody>
</table>

The two measures are clearly inversely related: as the proportion of the data accounted for by the four most frequent infinitives is decreasing across the four centuries, the type/token ratio is consistently increasing. This suggests two things about the development of the [tener que + Infinitive] construction. First, the slot occupied in the construction by an infinitive is becoming more schematic (Bybee 2010:25). That is, the slot is no longer limited to the set of highly frequent infinitives, but instead can be occupied by a variety of infinitives.

In the 15th century, the four highly frequent infinitives accounted for 64% of the data. The construction’s occurrence at that point was likely linked to the fact that those four verbs are generally quite frequent and that each of them contributes a sense of purpose, intention, and/or responsibility to the construction. By the 18th century, those four infinitives only account for 24% of the data, which now include multiple tokens of infinitives such as temer ‘to fear’, sufrir ‘to suffer’, and venir ‘to come’, which would be semantically incongruous with the original lexical meaning of [tener que + Infinitive], the possession of a concrete object designated as the patient of the infinitive’s action. The examples in (100) and (101) illustrate this innovative usage.
(100) **el otro día tenía que venir a cantar una misa**

the other day (he) **had to come** to say a mass [18th c.; *DLNE*: 631]

(101) **he tenido que sufrir** (gracias a la recomendación de usted) casi todo el primer acto

I have had to suffer (thanks to your recommendation) almost the entire first act [18th c.; *CV*: 115]

The second generalization we can make based on Table 3-5 is that the construction is also becoming more productive, i.e. more likely to occur with new infinitives. From the 15th to the 16th century, the number of distinct infinitive types occurring in the target construction quadruples from 26 to 105. In Section 3.1.2, it was noted that the transition between these two centuries also marks a clear jump in the overall frequency of the construction. The number of infinitives increases for the 17th century (to 122) and again for the 18th (to 165), meaning that by the end of this period of four centuries, the construction occurs with 139 more infinitives than in the 15th century data.

These two changes in the types of infinitives and their distribution are closely linked to the evolving semantic status of [*tener que* + Infinitive], as Bybee states that highly grammaticalized constructions are both highly schematic and highly productive (2010:94). As [*tener que* + Infinitive] begins to be analyzed as containing less of its original lexical meaning and instead as a modal construction denoting obligation, the constraints on its distribution throughout the grammar are increasingly relaxed. In the 15th century, evidence of those constraints that might be linked to the lexical history of *tener* ‘to have’ is visible (the notion of Persistence proposed by Hopper). But by the end of the time period investigated here, those constraints appear to have been all but removed, with [*tener que* + Infinitive] occurring with more and more infinitives. Furthermore, a number of these are only available in an occurrence of the
construction for which the interpretation of tener que is more grammatical (obligation) than lexical (possession).

3.3. Structural changes in [tener que + Infinitive] evident in the 15th-18th centuries

Evidence of the ongoing grammaticalization of [tener que + Infinitive] discussed to this point has been largely semantic in nature. In addition to the construction’s content, its formal components also exhibit changes in this time period. The primary source of evidence of structural changes comes from the placement of a syntactic object in the construction. As previously mentioned, the origin of the grammaticalizing construction is the recurring statement of possession of an overt object designated for a specific purpose. This object may intervene between tener and que, yielding the structural arrangement [tener (X) que + Infinitive]:

(102)  no tengo cerca dello otra cosa que dezir
       ‘I do not have anything else to say about that’ [16th c.; DLNE: 95]

As the construction increases in frequency and undergoes the semantic changes discussed in Section 3.2, tokens with frequent infinitives begin to take on conventionalized meanings, especially with hacer ‘to do’ and ver ‘to see’ (3.2.3 and 3.2.4 respectively). In these instances, the relative pronoun que ‘which’ often serves as the object itself, with no other expressed object:

(103)  como quien no tiene qué perder
       ‘like someone who does not have (anything) to lose’ [15th c.; Celestina: 338]
Tokens with *que* as the object are semantically identical to those with explicit objects. While unspecified, the object is still present and can therefore be interpreted as the syntactic object of *tener*, and the semantic patient of the infinitive’s action.

In contrast, tokens for which the object does not intervene between *tener* and *que*, but instead is postposed to the infinitive, begin to appear and increase in frequency from the 15th to the 18th century. For these tokens, it is more difficult to interpret the object as that of *tener*, and instead they suggest that it is the object of the infinitive verb.

(104) *para descargo de su conciencia tiene que hacer una denuncia*

‘to relieve his conscience he **has to make** a report’ [18th c.; DLNE: 573]

Only one token with a postposed object as in (104) appears in the 15th century data for the combined corpora. However, this structural type makes up a consistently larger portion of the data in each subsequent century, eventually rising as high as 25% in the 18th century as depicted in Table 3-6.

Table 3-6. Percentage of tokens of [*tener que* + Infinitive] with postposed objects from the 15th to the 18th century

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
</tr>
</thead>
<tbody>
<tr>
<td>% data exhibiting postposed object (N/total tokens)</td>
<td>1% (1/81)</td>
<td>9% (27/316)</td>
<td>16% (53/327)</td>
<td>25% (102/404)</td>
</tr>
</tbody>
</table>

This arrangement of the elements in the construction (with the object postposed to the infinitive verb) is congruous with the grammaticalized modal construction as it commonly appears in Modern Spanish, yielding an unambiguous obligation meaning.

That change parallels the advent of tokens for which there is no object present, i.e. the infinitive is intransitive:
The use of an intransitive infinitive would be incompatible with the lexical meaning of the construction (possession of an object), but not with obligation. For that reason, the advent of tokens lacking an object is taken as additional evidence for the semantic evolution of \([tener que + \text{Infinitive}]\). This also has implications for the status of \(que\), which is not pronominal in (104) or (105). It is lacking any anaphoric reference in obligation uses of the construction where there is no possessed object to which it might refer.

Examples like (105) illustrate that as \(que\) has been depleted of its pronominal meaning, it has effectively become fused with \(tener\) in the \([tener que + \text{Infinitive}]\) construction. This type of change, which has been termed “depronominalization” (Torres Cacoullos 2002) not only bleaches the erstwhile relative pronoun of its semantic content but also facilitates its incorporation into the construction to form a single unit: \(tener que\). Heine and Kuteva (2002: 254) also list several unrelated languages in which relative pronouns have become complementizers, suggesting that the present change concerning the relative pronoun \(que\) is not altogether unusual. In this case, the \(tener que\) construction does not appear without \(que\), but as can be seen in modal tokens like (105), when the lexical verb (e.g. \(hablar\) ‘to talk’) does not carry with it a direct object to which \(que\) would refer, \(que\) cannot be considered pronominal.

With the change in the status of \(que\) in mind, Figure 3-5 illustrates the three clear trends in the placement (or absence) of an object in the construction across this time period.
As can be seen, the two object types associated with the modal meaning of [tener que + Infinitive] increase steadily from the 15th to the 18th century. Tokens without an object mirror the gains made by those with postposed objects, increasing from 5% (4/78 tokens) of the 15th century data to 29% (119/404) of the 18th century data. Conversely, the use of que as a pronominal object plummets from 59% (46/78 tokens) in the 15th century to just 6% (27/404) in the 18th.

In sum, the two types of object placement associated with the modal meaning of [tener que + Infinitive] (postposed objects and object-less tokens) both exhibit steady increases in frequency. The use of que as a pronominal object per se, which is associated with the lexical (possession) meaning of the construction, diminishes from making up more than half of the data at the beginning of this time period to accounting for only about a tenth of that amount by the end. It is tempting, therefore, to assume that this indicates a clear loss of analyzability by the construction; that is, the degree to which tener and que are conceived of as separable components...
of the structure (Langacker 1987: 292). If an object never intervenes between tener and que, they might be considered to be a single unit rather than two elements.

It is important to note, however, that tokens with overt objects intervening between tener and que do not disappear from use. These tokens preserve the older relative clause-type (possession) form of the construction and continue to make up a substantial portion of the data, never accounting for less than 18% of the data in any of the four centuries discussed here. But in the 19th and 20th century data (discussed in Chapters 4 and 5, respectively) this portion diminishes to 2%, and then appears to have bottomed out in the modern language (3% in the 20th century), as depicted in Table 3-7.

Table 3-7. Percentage of tokens of [tener que + Infinitive] with intervening objects from the 15th to the 18th century

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
<th>19th</th>
<th>20th</th>
</tr>
</thead>
<tbody>
<tr>
<td>% data exhibiting intervening object (N/total tokens)</td>
<td>19% (15/81)</td>
<td>38% (121/316)</td>
<td>29% (95/327)</td>
<td>18% (71/404)</td>
<td>2% (13/565)</td>
<td>3% (36/1149)</td>
</tr>
</tbody>
</table>

In the 19th and 20th centuries, although relative clause-type tokens appear, the majority of them are examples of tener que ver (see Section 3.2.4), and tokens for which an object intervenes between tener and que are exceedingly rare26. For this reason, the following two chapters will focus on modal uses of the construction and take as a given that que no longer represents the pronominal category it once did, having fused together with tener as a single unit in the modal construction.

26 A considerable number of the tener que ver tokens in the 19th and 20th centuries are negated, as in el parto no tiene nada que ver con ellos ‘birth doesn’t have anything to do with them (men)’ [20th c.; MexCult]. Although such tokens ostensibly include an object (e.g. nada ‘nothing’) intervening between tener and que, this is evidently a conventionalized usage that is considered here to be distinct from the older possession construction.
For the present purposes, the persistence of tokens with overt objects intervening between tener and que in the 18th century data affirms that the general [tener que + Infinitive] construction did not grammaticalize wholesale during this period. The older form of the construction with the original lexical meaning of tener ‘to have’ persists in the 18th century (and occasionally appears in the 19th and 20th centuries):

(106) en el asumpto no sabía más ni tenía otra cosa especial que desir.

‘On the matter, he did not know nor did he have any other thing to say’

[18th c.; DLNE: 572]

Still, these are specific uses of the construction, and by the 18th century they are far less frequent than tokens indicating obligation. In the combined 18th century data, 231 of the 404 tokens in the 18th century either have a postposed object or no object. That means that 54% of the total data for the latest century are associated with the more modal meaning of the construction. In contrast, only 98 tokens (24%) have either an intervening object or exhibit a pronominal use of que, the two types of object placement more closely associated with the possession meaning of [tener que + Infinitive].

Also, the tokens with pronominal que objects and intervening objects are often tokens with one of the four highly frequent infinitives discussed in 3.2.6 (dar ‘to give’, decir ‘to say’, hacer ‘to do’ and ver ‘to see’). While these infinitives account for only 26% of the overall data in the 18th century, they make up 41% of the tokens with que objects and intervening objects (40 of 98 tokens). This is worth mentioning, as Bybee notes that “types of extremely high token

27 A number of tokens in the combined corpora have an overt object placed in front of tener, preposed to the rest of the construction: Esto es lo que tenemos que certificar ‘this is what we have to certify’ [18th c.; DLNE: 555]. These are often ambiguous and are therefore not counted as necessarily being more closely associated with either the lexical or the modal meaning of [tener que + Infinitive]. This is why the counts for the other four types of object placement do not add up to 100% of the total number of tokens in the 18th century.
frequency contribute less to productivity because of their autonomy and loss of analyzability” (2010:95). With this in mind, it appears that conventionalized instances of the construction with frequent infinitives, such as tener que hacer ‘to have (something) to do’ are relegated to their own specific niche in the semantic web occupied by the general [tener que + Infinitive] construction. That is, they are conservative uses, effectively entrenched in an older kind of manifestation of the tener que construction.

Highly frequent infinitives like hacer ‘to do’ would then be less of a factor in driving the productivity of the grammaticalizing construction than are instances for which the construction is not associated with a conventional possession sense. The ‘have (something) to do’-type uses are entrenched, in a sense, in the realm of interpretations for which it is conventionally understood that at least some sense of possession is intended. They may have been crucial at first, in creating the context for inferring obligation, but are evidently less of a factor in spreading the construction to new infinitive types. Again, this leads to the conclusion that tokens with postposed objects and tokens without objects (intransitive infinitives) are the locus of productivity for the construction, as they are more closely associated with the innovative obligation meaning.

3.4 Conclusion

The period from the 15th to the 18th century is a crucial juncture in the ongoing grammaticalization of [tener que + Infinitive]. A substantial increase in the overall frequency of the construction is evident as it goes from being nearly nonexistent in the 14th century to being a conventionalized means of expressing obligation in the 18th. This rising rate of occurrence is undoubtedly linked with the semantic changes also taking place at this time. During these four

28 This finding is parallel to that of Bybee and Torres Cacoullos (2009): lexically-specific instances of the Spanish progressive construction featuring lexical verbs that were not associated with other semantically similar verbs did not contribute to the productivity of the general construction (e.g. estar esperando ‘to be waiting’, where esperar ‘to wait’ is not part of a set of related ‘waiting’ verbs).
centuries, tokens begin to appear for which the meaning of [*tener que* + Infinitive] is semantically-structurally ambiguous, simultaneously denoting possession and a sense of responsibility. The ambiguity here is perhaps best considered “indeterminacy” in that it arises in contexts for which a speaker and hearer need not have exactly the same meaning in mind (see Section 3.2.2) and has, in fact, been argued to facilitate grammaticalization (Schwenter and Torres Cacoullos 2008; Croft 2010). This conceptualization of ambiguity contrasts with a definition of ambiguity based on distinct, mutually exclusive parsings (presumed underlying syntactic structures) of a sentence. By the end of the time period considered here, examples of the *tener que* construction that allow only an interpretation of obligation are common.

An advantage of analyzing the semantic and functional evolution of the erstwhile possession construction via a diachronic text-based investigation is that it affords insight into the mechanisms by which this change takes place. The rising frequency of [*tener que* + Infinitive] parallels the construction’s extension to a steadily increasing number of distinct infinitive types. Originally confined to a limited set of infinitives, the construction is generalized to more and more infinitives with each subsequent century. At the same time, the set of specific, highly frequent infinitives that accounted for the majority the data in the 15th century makes up fewer and fewer of the tokens in the following centuries.

The quantitative evidence presented here also illustrates the structural changes that take place within the [*tener que* + Infinitive] construction as its modal meaning becomes more conventionalized. Arrangements of the construction’s components that reflect its modern modal usage become more frequent, while arrangements associated with its original meaning become more limited. Despite the potential ambiguity of many tokens during this time period, objective measures such as object placement allow for making the generalization that obligation uses of the construction are on the rise.
That is not to say, however, that the advent of innovative modal uses of [tener que + Infinitive] coincide with a complete disappearance of lexical possession tokens. While the latter are more limited in their frequency and evidently in the range of infinitives with which they occur, the use of [tener que + Infinitive] to encode possession continues to be viable, even in the 18th century. With this in mind, the persistence of older meanings is perhaps just as important as the appearance of newer meanings to a discussion of the grammaticalization path of this construction. The semantic and formal changes it undergoes during this period prove to be gradual and variable: new uses coexist with old ones, and the leap from possession to obligation is in no way a sudden transition. From a theoretical perspective, this evidence provides support for predictions made in the existing grammaticalization literature regarding gradualness, layering, and variability (Hopper and Traugott 1993: 35-37; Bybee 2010: 114-119).

The analysis done to this point has focused on the grammaticalization of the [tener que + Infinitive] construction itself. But in addition to variation within the construction, there is variation without, as several other periphrastic modal constructions overlap with [tener que + Infinitive] in the functional domain of expressing obligation. The following chapter will discuss the variation between these alternative variants (haber que ‘to be necessary to’, haber de ‘must’, deber (de) ‘should, ought to’) and the construction analyzed here.

Each of these constructions is itself a grammaticalizing reflex of an earlier lexical verb and as such, displays a range of different meanings as does [tener que + Infinitive]. Semantic and functional variation both within and between the alternative variants is a focal point of the analysis to follow, as it pertains to the ongoing grammaticalization of [tener que + Infinitive] itself. Analyzing that variation will draw from the tendencies presented here, considering the apparent importance of quantifiable and qualifiable factors such as the placement of a syntactic object and the infinitive verb type appearing in the construction. This variation, as it proceeds in the 19th and 20th centuries, provides additional evidence of the mechanisms and processes at work
as the construction grammaticalizes, and allows for refining the generalizations and conclusions drawn here.
Chapter 4

Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 19th century

The previous two chapters have analyzed the internal evolution of [tener que + Infinitive] as it coalesced into an expression of obligation. By the 18th century, the construction has acquired its modern sense of obligation and has been generalized for use with a broad range of infinitives. But its coexistence and variation with other periphrastic constructions is equally important to charting the route it took from a particular usage of a possession verb to a modal construction. Along with [tener que + Infinitive], modern reference grammars of Spanish list several other constructions with modal auxiliary verbs that may be used to indicate obligation. Specifically in combination with an infinitive, deber (de) ‘should, ought to’, haber que ‘to be necessary to’, and haber de ‘must’ can, depending on context, be applied to the expression of obligation (Butt and Benjamin 2004: 327-330).

To characterize the variation between these constructions, the following sections outline a variationist analysis of the variants. First, the constructions themselves are described, and their functional equivalence is established. Their relative frequencies for the centuries leading up to and including the 18th are then determined. Finally, multivariate analyses are presented, describing the factors that influence the occurrence of [tener que + Infinitive] as one of several possible variants for the expression of obligation in Spanish in the 19th century. Describing and characterizing the variation among these modal constructions will aid in determining the status and position of [tener que + Infinitive] in the modal system and can effectively be taken as an indication of the relative degree of grammaticalization of the construction.

29 Several other constructions are attested in the written record, including [tener de + Infinitive] which is discussed in Chapter 2. The four constructions listed here are the forms that survive in the modern language and continue to occur with substantial frequency. As such, they are the focus of the present analysis.
4.1. Alternative modal verb constructions for the expression of obligation

The four constructions commonly put forward as periphrastic expressions of obligation all share structural characteristics in addition to their semantic similarity. The examples in (107)-(110) depict the usage of these four constructions to convey a sense of obligation.

*haber de* + Infinitive

(107) *para pagarselas a dicho alcalde mayor, ubo de*

for-to pay.INF.him.them to said magistrate senior have-to.PST.3SG

*comprar* media arroba

buy.INF half arroba

‘in order to repay the senior magistrate, (he) had to buy a half arroba’

[18th c.; DLNE: 452]

*haber que* + Infinitive

(108) *Siéntese usted…Y no hay que asustarse ni alborotarse [...] por nada de lo-que yo diga*

Sit.IMP you and NEG exist.PR.SG REL frightening.REFL.INF nor upset.REFL.INF [...] by nothing of that-which I say.PR.SBJV

‘Sit down…and you need not be frightened or upset by any of what I say’

[18th c.; Si: 275]

---

30 The arroba is a historical Spanish weight measure
deber (de) + Infinitive

(109) Pero DEBE usted serenarse, y esperar que la suerte mude nuestra aflicción presente

But should.PRS.3SG you calm-down.INF and hope.INF COMP DEF luck change.PRS.SBJV our affliction present

‘You should calm down and hope that luck changes our present affliction’

[18th c.; Sí: 222]

tener que + Infinitive

(110) pero lo cierto es que el sujeto tendrá que contentarse con sus quince doblones

but that-which certain be.PRS.3SG COMP DEF subject have-to.FUT.3SG be-content.INF with his fifteen doubloons

‘but what is certain is that the individual will have to be content with his fifteen doubloons’ [18th c.; CN: 77]

In each case, the modal verb (deber, haber or tener) is followed by an infinitive lexical verb. Between the two verbs may intervene either que or de. For the construction with deber, the appearance of de is variable (see Section 4.1.3). As with the erstwhile relative pronoun que appearing in [tener que + Infinitive], the intervening element contributes no semantic content to the construction (Section 3.3). In examples (107) – (110), each of which is from the 18th century, it is clear that all four of the modal verb constructions can be applied to the expression of obligation. Still, their meanings are often nuanced and depend largely on the context in which they appear, as has been discussed for [tener que + Infinitive] in Chapter 3. It is therefore
worthwhile to review each of the other three constructions, as their origins and usage relate closely to \[tener que + Infinitive\].

### 4.1.1. haber que + Infinitive

Of the three alternative modal constructions, \[haber que + Infinitive\] is the most closely related to \[tener que + Infinitive\] in its structure. Based on a possession verb, \textit{haber} ‘to have’, this periphrastic construction already appears in the 12\textsuperscript{th} century in the collected corpus used in Chapters 2 and 3. In her treatment of Spanish modal verb constructions, Yllera also places the initial appearance of \[haber que + Infinitive\] in that epoch (1980: 109). She notes that its genesis is analogical to that of \[tener que + Infinitive\], characterized by tokens in which \textit{que} ‘something’ is a possessed object upon which the action of the included infinitive is performed:

\begin{align*}
\text{(111)} & \quad \text{ellos no} \quad \text{an} \quad \text{que} \quad \text{comer} \quad \text{n}n \quad \text{traen} \quad \text{armas} \\
& \quad \text{they NEG have.PRS.3PL that-which eat.INF nor carry.PRS.3PL weapons} \\
& \quad \text{‘they do not have anything to eat nor carry any weapons’ [13\textsuperscript{th} c.; GEI: II, 134]}
\end{align*}

Furthermore, there is apparent congruity between the components of each of the two constructions. In addition to both being based on possession verbs that have incidentally alternated in other functions (see Chapter 2), the element \textit{que} was originally a relative pronoun in \[haber que + Infinitive\] as it was with \[tener que + Infinitive\]. However, tokens like that in (112) suggest that the semantic development of modal meaning by \[haber que + Infinitive\] was not necessarily identical to that of \[tener que + Infinitive\].
As discussed in Chapter 3, a number of tokens of \( \text{tener que} + \text{Infinitive} \) appear in this exact context, in reference to an official document or proceeding that was reviewed and approved by a magistrate, who determined that he had nothing to add to it. However, the manifestation of \( \text{haber} \) in (112) is a conventionalized combination of \( \text{haber} \) with the adverb \( y \) ‘there’ (from Latin \( \text{ibi} \) ‘there’). This combination eventually coalesced as a single word \( \text{hay} \) (or \( \text{ay} \) as in the example above) (Penny 2002:191). That is to say that the eventual modal sense of the construction with \( \text{haber} \) may be based on the use of the verb to indicate existence (112) rather than possession (111).

This is the modern meaning of \( \text{hay} \) ‘there is’, so the usage of \( \text{[haber que + Infinitive]} \) in (112) suggests that the meaning is closer to ‘there \text{was} nothing to add’ than to ‘he \text{had} nothing to add’, the meaning that similar tokens of \( \text{tener que} + \text{Infinitive} \) would have. This change from expressing possession to expressing existence with \( \text{haber} \) could explain why tokens of \( \text{tener que} + \text{Infinitive} \) that clearly convey obligation seem to predate such tokens for \( \text{haber que} + \text{Infinitive} \) in the corpus utilized here\(^{31}\). The delayed appearance of obligation tokens with \( \text{haber} \) could explain why tokens of \( \text{tener que} + \text{Infinitive} \) that clearly convey obligation seem to predate such tokens for \( \text{haber que} + \text{Infinitive} \) in the corpus utilized here\(^{31}\).

\(^{31}\) Yllera presents tokens of \( \text{haber que} + \text{Infinitive} \) without the adverbial element \( y \) ‘there’ that appear to be synonymous with the example in (6), e.g. \( \text{que mucho a que emendar} \) ‘(he) \text{has} much to add’ (1980: 109). Such occurrences would suggest that the evolution of this construction did not necessarily pass through a stage in which it expressed existence rather than possession.
que + Infinitive] would otherwise be surprising, as haber predates tener in numerous other contexts as the predominant possession verb in Spanish (see Chapter 2).

Despite this difference, the eventual transition to expressing obligation is similar to that of [tener que + Infinitive]. The included sense of necessity or responsibility in contexts such as ‘there was nothing to add’ leads to the inference that nothing needs to be added, and from there to a more general sense of obligation. This semantic transition through ambiguity is depicted in (113) – (114), both from the 18th century.

(113) ¿quién puede oponérsele? Nada hay que temer.

‘Who can disagree with that? There is nothing to fear.’ [18th c.; St: 223]

(114) Bien sabe Dios que no tengo la culpa...Venga usted aquí. No hay que temer.

‘God knows it is not my fault…come here. There is nothing to fear.’
[18th c.; St: 280]

In (113-114), the use of the haber que construction occupies a middle ground between objectively denying the existence of something to be feared (nada ‘nothing’ in (113)) and not needing or being obligated to fear something. The example in (114) is perhaps closer to the latter sense; without the overt object that appears in (113), the sense of obligation is foregrounded, allowing for the inference that rather than ‘you have nothing to fear’, the speaker is suggesting ‘you do not have to be afraid’. This leads to uses where the implication of obligation is more direct, as in (108) above, as with the semantic change undergone by [tener que + Infinitive] described in Chapter 3. Still, the form of haber that appears in the haber que construction is invariably with a third person subject, and typically in the present tense hay ‘there is’ or ‘there are’. This also supports the notion that the evolution of this construction is not entirely analogous to that of [tener que + Infinitive].
4.1.2. *haber de* + Infinitive

Despite its apparent etymological similarity to [*haber que* + Infinitive], the alternative construction [*haber de* + Infinitive] has a distinct history. Yllera asserts that already in Vulgar Latin, the preposition *de* ‘of, from’ was used to introduce an infinitive, and that in Medieval Spanish, the combination [noun + *de* + infinitive] was used to express obligation or possibility: e.g. *fijas de casar* ‘daughters to marry’ from the 12th century *Cantar de Mío Cid* (1980: 96). The *haber de* construction appears in the present corpus as early as the 12th century and already in the earliest tokens it is used to express obligation:

\[(115) \text{et por cuyo consejo fazían lo-que avían de fazer} \]
\[
\text{and by POSS advice do.IPFV.3PL that-which have.IPFV.3PL of do.INF}
\]

‘and by his advice they did what they had to do’ [13th c.; *Calila*: 225]

It appears that there is no evidence of the *haber de* construction passing through an intermediate phase of ambiguity with implied obligation in Medieval Spanish as is the case for *haber que* and *tener que*. Rather, it had already acquired modal meaning by the time the latter two were beginning to appear with their original lexical meaning. Given the several centuries’ head start enjoyed by [*haber de* + Infinitive], it is perhaps unsurprising that it has evidently progressed even further in its grammaticalization and acquisition of modal meaning. In fact, in Modern Spanish the meaning most commonly associated with this construction is one of probability, or epistemic modality, rather than obligation (Butt and Benjamin 2004: 329-330). This is evident in examples like (116).
(116) *Búscale bien, que por ahí ha de estar.*

‘Look closely, it **must be** there.’ [18th c.; Sí: 252]

Probability is an attested target meaning acquired by grammaticalizing forms that first express obligation (Coates 1983; Bybee, Perkins and Pagliuca 1994: 195; Heine & Kuteva 2002:218). In fact, probability can be conceived of as a form of obligation, or a consequence of external or pragmatic obligation (Gómez Torrego 1999: 3353). This is the case for (116), where the speaker’s world knowledge presumably requires that the object be located where he expects it to be. Butt and Benjamin use this very context to explain that [*tener que* + Infinitive] may also be used to “indicate a strong supposition, as in *búscalo bien, tiene que estar ahí* ‘check thoroughly, it’s **got to** be there’” (2004: 328).

4.1.3. *deber (de) + Infinitive*

The third construction that commonly appears in the corpus analyzed in Chapters 2 and 3 for the expression of obligation is [*deber (de) + Infinitive*]. *Deber* ‘should, ought to’ is derived from the semantic source ‘to owe’, a grammaticalization path attested for this source-construction meaning in numerous languages including English *ought* (Denning 1987; Traugott and Dascher 2002; Heine and Kuteva 2002: 227). Deber retains its original meaning when used as a lexical verb in Modern Spanish, but when used in conjunction with an infinitive indicates an obligation to perform the action of that infinitive. The earliest occurrences of *deber* in the present corpus convey that sense:

---

32 Incidentally, the etymological source of *deber* ‘to owe’ is the Latin *dēbeo* ‘I owe’, which itself is a shortening of *dehabeo*, a conventionalized form of uses like *de aliis habeo* ‘I have that which is another’s’ (Vaan 2008: 162).
(117) *Et dizen que el omne entendido deve contar a su padre et a su madre por amigos*

‘and they say that the wise man *should count* his father and mother as friends’

[13th c., *Calila*: 274]

A particular type of occurrence of the *deber* construction includes the preposition *de* ‘of, from’, which intervenes between the auxiliary verb and the infinitive, as it does with [*haber de* + Infinitive]. It is commonly argued in prescriptive grammars that the inclusion of *de* is not (or should not be) for use with expressing obligation, but instead for denoting probability, as with the examples in (118-119).

(118) *Por cierto mal te trataba ese hombre. Mala gente debe de ser recueros.*

Certainly that man mistreated you. Mule drivers *must be* bad people.

[16th c.; *Crotalón*: Cuarto canto]

(119) *Pues agora calla, que llaman a la puerta, que deben de venir a comprar.*

Be quiet now, they are calling at the door; they *must be coming* to shop.

[16th c.; *Crotalón*: Cuarto canto]

However, Butt and Benjamin concede that “this mistake is not uncommon in popular speech […] and even in writing (2004:328). There is also descriptive evidence confirming the use of *deber de* to express obligation. Yllera argues that the distinction between the two *deber* constructions (with and without *de*) is only artificial (1980:128). In a study of Mexican Spanish, Fairclough finds evidence of a ‘total semantic fusion between the forms *deber* and *deber de* (2000:1). Balasch (2008) corroborates the fusion of the two forms with respect to expressing obligation via a diachronic text-based analysis, and Blas Arroyo (2011) provides further evidence of the same from both Latin American and Iberian varieties of spoken Spanish. For these reasons, the two forms with *deber* are treated in the present analysis as a single construction: [*deber (de) +
Infinitive]. The functional equivalence of the deber de construction with [tener que + Infinitive] is also evident in examples like (120) and (121), from the 16\textsuperscript{th} and 20\textsuperscript{th} centuries respectively.

(120) \ldots y entendía que para yo no ser tan asno como mi amo que debía de saber algún latín.

‘…and I understood that in order to not be as much of an ass as my master, that I ought to know some Latin.’ [16th c.; Crotalón: Cuarto canto]

(121) …porque tiene uno que comprender...como el marido debe de comprender a la mujer…

‘…because one has to understand…like a husband ought to understand his wife…’ [20\textsuperscript{th} c.; MexPop: 317]

4.2. Variation among the four alternative modal constructions: 12\textsuperscript{th}-18\textsuperscript{th} centuries

In the corpus utilized for the discussion of [tener que + Infinitive] from the 12\textsuperscript{th} to the 18\textsuperscript{th} centuries, each of the modal constructions described in Section 4.1 is attested. To determine the relative frequencies with which the variants were used to encode obligation over that time period, tokens of each were extracted from the same corpus of texts used in Chapters 2 and 3. Those that expressed obligation (as opposed to possession or probability) were counted for each century, as depicted in Table 4-1.

Table 4-1. Token counts for constructions encoding obligation for the 12\textsuperscript{th}-18\textsuperscript{th} centuries

<table>
<thead>
<tr>
<th>Century</th>
<th>12\textsuperscript{th}</th>
<th>13\textsuperscript{th}</th>
<th>14\textsuperscript{th}</th>
<th>15\textsuperscript{th}</th>
<th>16\textsuperscript{th}</th>
<th>17\textsuperscript{th}</th>
<th>18\textsuperscript{th}</th>
</tr>
</thead>
<tbody>
<tr>
<td>[deber (de) + Inf.]</td>
<td>2</td>
<td>405</td>
<td>210</td>
<td>137</td>
<td>132</td>
<td>125</td>
<td>45</td>
</tr>
<tr>
<td>[haber de + Inf.]</td>
<td>12</td>
<td>197</td>
<td>119</td>
<td>181</td>
<td>172</td>
<td>336</td>
<td>111</td>
</tr>
<tr>
<td>[haber que + Inf.]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>[tener que + Inf.]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>
This extraction was not from the entire 12th – 18th century corpus: data were not extracted from three texts (Estoria de España, Libro de Buen Amor, and Segunda comedia de Celestina from the 13th, 14th, and 16th centuries respectively) because they were not needed to establish the basic pattern of occurrence for the constructions. Until the 17th century, [tener que + Infinitive] and [haber que + Infinitive] are effectively nonexistent as expressions of obligation, so the haber de and deber (de) constructions are clearly the predominant forms. At least two documents in each century from the 13th to the 16th were analyzed, such that at least 100,000 words were included in the sample for each century. This is with the exception of the 12th century, for which the only document is Cantar de mio Cid, a text of just over 30,000 words. The pattern of usage for the variants, displayed in Figure 4-1, was immediately apparent from this sample.

Figure 4-1. Percentages of data accounted for by four modal constructions expressing obligation from the 12th to the 18th century

---

33 The Cantar de Mio Cid has an unusually high number of tokens of [haber de + Infinitive] compared to the percentage of the data composed of that variant for the following two centuries. Yllera also notes the high frequency of [haber de + Infinitive] in her counts (1980: 98).
As noted, \([\text{deber (de)} + \text{Infinitive}]\) and \([\text{haber de} + \text{Infinitive}]\) are by far the predominant constructions throughout this seven-century time period. However, there are two important trends that bear upon the analysis of \([\text{tener que} + \text{Infinitive}]\). The first of these is that the portion of the data accounted for by the \(\text{deber (de)}\) construction diminishes with each passing century. The second trend is that the \(\text{haber de}\) construction undergoes proportional increases in relative frequency as if encroaching on the domain of \([\text{deber (de)} + \text{Infinitive}]\). However, the 17th and 18th centuries present an important departure from this pattern: the \(\text{haber que}\) and \(\text{tener que}\) constructions go from being negligible in the 17th century to making up a combined 14% of the 18th century data. In this century, while \([\text{deber (de)} + \text{Infinitive}]\) continues to decrease, so does \([\text{haber de} + \text{Infinitive}]\) for the first time. The 18th century data are quite suggestive of what is to come in the 19th century as the modern pattern of variation among these constructions continues to crystallize. This is the subject of the following sections, in which \([\text{tener que} + \text{Infinitive}]\) is analyzed alongside the other variants as an emerging modal verb construction.

### 4.3. Distribution of data in the 19th century

To assess the status of the \([\text{tener que} + \text{Infinitive}]\) construction in the 19th century, a selection of written texts was compiled, from which were extracted all tokens of the four modal verb constructions discussed in Section 4.1. This 19th century corpus includes 4 novels and 7 theatrical plays (all of which are written in prose and represent peninsular Spanish), as well as the \textit{Documentos Lingüísticos de la Nueva España} (Company Company 1994), representing Mexican Spanish. To these sources were added supplemental tokens from Davies’ (2002-) online \textit{Corpus del Español}. Table 4-2 outlines the corpus.
Table 4-2. Corpus of 19th century texts

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novels</strong></td>
<td></td>
</tr>
<tr>
<td>La Regenta (Tomo I)</td>
<td>146,000</td>
</tr>
<tr>
<td>Doña Perfecta</td>
<td>65,500</td>
</tr>
<tr>
<td>Los pazos de Ulloa</td>
<td>83,800</td>
</tr>
<tr>
<td>Pepita Jiménez</td>
<td>56,500</td>
</tr>
<tr>
<td><strong>Novels total</strong></td>
<td><strong>351,800</strong></td>
</tr>
<tr>
<td><strong>Theater</strong></td>
<td></td>
</tr>
<tr>
<td>Acertar errando</td>
<td>14,000</td>
</tr>
<tr>
<td>La gran comedia</td>
<td>16,000</td>
</tr>
<tr>
<td>Los presupuestos</td>
<td>16,000</td>
</tr>
<tr>
<td>Amor de padre</td>
<td>18,700</td>
</tr>
<tr>
<td>¡El siete!</td>
<td>5,000</td>
</tr>
<tr>
<td>No hay mal que por bien no venga</td>
<td>15,500</td>
</tr>
<tr>
<td>Las circunstancias</td>
<td>15,000</td>
</tr>
<tr>
<td><strong>Theater total</strong></td>
<td><strong>100,200</strong></td>
</tr>
<tr>
<td><strong>Documentos Lingüísticos de la Nueva España</strong></td>
<td>21,000</td>
</tr>
<tr>
<td><strong>Document total</strong></td>
<td><strong>21,000</strong></td>
</tr>
<tr>
<td><strong>Supplemental texts</strong> (Davies 2002-)</td>
<td></td>
</tr>
<tr>
<td>El antecristo</td>
<td>75,180</td>
</tr>
<tr>
<td>La Princesa de Viana</td>
<td>36,748</td>
</tr>
<tr>
<td>Compendio de la vida de San Alfonso María de Ligorio</td>
<td>17,509</td>
</tr>
<tr>
<td>La mujer de Navarra</td>
<td>6309</td>
</tr>
<tr>
<td>Doña Blanca de Navarra, crónica del Siglo XV: intitulada Quince días de reinado</td>
<td>159,114</td>
</tr>
<tr>
<td>El caballero sin nombre</td>
<td>23,500</td>
</tr>
<tr>
<td>El copo de nieve</td>
<td>78,261</td>
</tr>
<tr>
<td>La mujer de todo el mundo</td>
<td>37,843</td>
</tr>
<tr>
<td>Genio e ingenio del pueblo andaluz</td>
<td>99,104</td>
</tr>
<tr>
<td>Manual del viagero y guía de los forasteros en Valencia</td>
<td>64,884</td>
</tr>
<tr>
<td>Historia de muchos Pepes</td>
<td>49,769</td>
</tr>
<tr>
<td><strong>Supplemental total</strong></td>
<td><strong>648,221</strong></td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td><strong>1,121,221</strong></td>
</tr>
</tbody>
</table>

Those tokens that express obligation were counted to determine their overall token frequencies as well as the relative frequencies of the four constructions. Table 4-3 illustrates how the pattern of occurrence for these four constructions has changed noticeably from the 18th to the 19th century.
Table 4-3. Token frequencies of periphrastic expressions of obligation in the 18th and 19th centuries (in number of tokens per million words)

<table>
<thead>
<tr>
<th>Century</th>
<th>18th N</th>
<th>18th Frequency</th>
<th>19th N</th>
<th>19th Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>[deber (de) + Inf.]</td>
<td>45</td>
<td>417</td>
<td>888</td>
<td>792</td>
</tr>
<tr>
<td>[haber de + Inf.]</td>
<td>111</td>
<td>1028</td>
<td>623</td>
<td>556</td>
</tr>
<tr>
<td>[haber que + Inf.]</td>
<td>11</td>
<td>102</td>
<td>172</td>
<td>153</td>
</tr>
<tr>
<td>[tener que + Inf.]</td>
<td>15</td>
<td>139</td>
<td>479</td>
<td>427</td>
</tr>
<tr>
<td>Total N</td>
<td>182</td>
<td></td>
<td>2162</td>
<td></td>
</tr>
<tr>
<td>Total # words</td>
<td>108,000</td>
<td></td>
<td>1,121,221</td>
<td></td>
</tr>
</tbody>
</table>

Immediately evident in Table 4-3 is the substantial increase in frequency by [tener que + Infinitive], more than tripling its overall frequency from 139 tokens per million words in the 18th century to 427 in the 19th century. It is worth noting that although these counts represent only those tokens that were interpreted as expressing obligation, lexical possession occurrences of the tener que construction persist in the 19th century data. In addition to the 427 obligation tokens, there were 39 occurrences of the relative clause construction that was described in Chapters 2 and 3 as the lexical source of the grammaticalizing usage. In addition, there were 11 tokens that were deemed to be expressions of probability rather than obligation, an apparently innovative use of the construction, as mentioned in Section 4.1.2.

This is taken as an indication that the innovative epistemic modal application of [tener que + Infinitive] is gaining a foothold in the 19th century but remains far less frequent than the deontic usage. Still, the recurrence of all three functions during this period provides quantitative evidence of Hopper’s principles of ‘layering’ and ‘persistence’ (1991:22). That is to say, grammaticalizing forms need not pass through discrete stages of development and categorically eliminate older, more conservative functions before acquiring new ones. Indeed, three developmental stages evidently coexist in this case.
The increase in frequency of obligation uses of \([tener\ que\ +\ \text{Infinitive]}\) is mirrored by \([deber\ (de)\ +\ \text{Infinitive}]\), and both of these changes are at the expense of \([haber\ de\ +\ \text{Infinitive}]\), as presented in Figure 4-2.

![Figure 4-2. Relative frequencies of modal constructions expressing obligation from the 17th to the 19th century](image)

While \([haber\ de\ +\ \text{Infinitive}]\) had accounted for 61% of the 18th century data, it only makes up 29% of the total combined tokens for the 19th century. The relative frequencies of both \([tener\ que\ +\ \text{Infinitive}]\) and \([deber\ (de)\ +\ \text{Infinitive}]\) expand proportionately to fill the space left by the receding \([haber\ de\ +\ \text{Infinitive}]\). The fourth construction, \([haber\ que\ +\ \text{Infinitive}]\), only increases slightly from 6% to 8% of the data.

It is now clear that rather than being dominated by just two constructions (those with \(haber\ de\) and \(deber\ (de))\) as it was from the 12th to the 18th century, this set of deontic modal verb constructions is characterized by a more even distribution of the four variants. In fact, \([tener\ que\]
+ Infinitive] is now almost as frequent as [haber de + Infinitive]. This change in the distribution is reflected in several instances in which tokens of distinct variants in the 19th century corpus are found in close contextual proximity.

**haber que and deber (de):**

(122) "Sí, sí; eso era él; y no había que hacerse ilusiones, ni buscar nueva manera de vivir. **Debía estar** satisfecho y lo estaba."

“Yes, yes, that was what he was, and he **should** not **entertain** any illusions or go in search of a new life. He **ought to** be satisfied and he was.”

[19th c.; Regenta: 470]

deber (de) and haber de:

(123) Allí no se debía alborotar porque al extremo de oriente, detrás de un majestuoso portier de terciopelo carmesí, estaba la sala del tresillo, que se llamaba el gabinete rojo. En este había de reinar el silencio, y si era posible también en la sala contigua.

“There **must** be no disturbance in this room because at its Eastern end, behind a majestic portière of crimson velvet, was the ombre-room, known as the red parlour. **Silence must** reign here and in the next room as well, if possible.”

[19th c.; Regenta: 250]

**haber que and haber de:**

(124) Lo que eran los hombres, y especialmente los indianos, lo que no les gustaba, la manera de marearlos, lo que había que conceder antes, lo que no se había de tolerar después...

“What men, particularly returned emigrants, were like, the things which displeased them, how to fluster them, what **had to be conceded** at first and what **should not be tolerated** afterwards…” [19th c.; Regenta: 230]

tener que and deber (de):

(125) No, yo tengo que seguir las huellas de María: **debo hacer** lo posible por encontrar la casa adonde nos han conducido

‘No, I **have to** follow Marías footprints: I **should** do everything possible to find the house to which they have taken us’ [19th c.; Davies 2002-]
The tokens in (122-124) illustrate that the variant constructions coincide in the 19th century texts and reinforce the assertion made in Section 4.1 that the four forms are employed as effectively functionally equivalent alternatives. The glosses for (122-124) are taken from the translation of *La Regenta* by Rutherford (1984). Note that in (123), Rutherford translates the alternating tokens of *[haber que + Infinitive]* and *[deber (de) + Infinitive]* using the same English modal verb, *must*.

Establishing the functional equivalence between these four constructions is a necessary component of the present analysis. It allows for comparing the distinct forms and assessing the extent to which specific linguistic factors, or contextual features, influence the choice of *[tener que + Infinitive]* over the three alternatives, thereby contributing to the ongoing grammaticalization of the target construction. This is especially important for a domain such as modality, in which meanings are said to be subtly nuanced (Coates 1983; Bybee, Perkins, and Pagliuca 1994).

However, examples like those in (122-125) provide support for the working hypothesis that all four of the constructions investigated here can, and often do, occupy the same functional space. Consider the numerous slightly different, but at least partially synonymous meanings of the English modals *must, should, ought to,* and *have to.* Sankoff refers to this type of situation in asserting:

…that some difference in connotation may, *upon reflection,* be postulated among so-called synonyms whether in isolation or in context, and that in the case of each one a number of competing syntactic constructions may be acceptable in different contexts, there is no reason to expect these differences to be pertinent every time one of the variant forms is used (1988b:153).

He goes on to state that when differences in grammatical function or meaning between forms are not always necessarily pertinent, these meanings can be ‘neutralized in discourse’, a mechanism
crucial to morphosyntactic variation and change (1988b: 153). With this in mind, the four modal constructions described in Section 4.1 are treated in the present analysis as alternating variants of a linguistic variable (Labov 1969; Sankoff 1988a).

4.4. Productivity of [tener que + Infinitive] in the 19th century

In Chapter 3, the number and type of infinitive verbs with which the tener que construction occurred was assessed as an indication of its ongoing grammaticalization. Furthermore, the type/token ratio, relating the number of distinct infinitive types to the total number of occurrences of the construction, was calculated for the 15th to the 18th centuries. Using this information, it was determined that as the construction increased in absolute frequency, it began to occur with a steadily increasing number of distinct lexical verb types. What is more, the portion of the tokens exhibiting one of the most frequent lexical types consistently decreased. That is, the tokens became more evenly distributed among the different infinitive types, with fewer and fewer lexical types recurring frequently. The data in the 19th century continue this trend of and increasing type/token ratio, as depicted in Table 4-4.

Table 4-4. Type/token ratios for lexical infinitive types appearing in [tener que + Infinitive] from the 15th to the 19th century

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
<th>19th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type/token ratio (# types/total N)</td>
<td>Type/token ratio (# types/total N)</td>
<td>Type/token ratio (# types/total N)</td>
<td>Type/token ratio (# types/total N)</td>
<td>Type/token ratio (# types/total N)</td>
</tr>
<tr>
<td></td>
<td>0.321 (26/81)</td>
<td>0.324 (105/324)</td>
<td>0.371 (122/329)</td>
<td>0.403 (165/409)</td>
<td>0.472 (226/479)</td>
</tr>
</tbody>
</table>

As indicated above in Table 4-3, the absolute frequency of [tener que + Infinitive] is more than three times higher in the 19th century than in the 18th. Concomitant with this change, the number
of distinct infinitives with which the construction occurs has increased from 165 to 226, raising the type/token ratio from 0.403 to 0.472. It is therefore evident that the construction has become substantially more productive, spreading throughout the grammar to appear with a diverse range of lexical verb types. This can be taken as evidence of a greater degree of grammaticalization of [tener que + Infinitive], as high productivity is considered a characteristic of highly grammaticalized constructions (Bybee 2010: 94).

Given this increase in productivity, it would appear that the 19th century data merely continue the trajectory indicated by the 15th-18th century data. However, a closer look at the lexical infinitive types appearing in the construction suggests that the explanation is more complicated. In section 3.2.6, it was noted that the four most frequent infinitive types appearing with [tener que + Infinitive] in the 15th to 18th centuries were hacer, ‘to do’, ver ‘to see’, decir ‘to say’, and dar ‘to give’. Of these, hacer, ver, and dar all could be argued to contribute an inherent sense of responsibility, while decir connotes at least an internally-imposed intention to say something on the part of the speaker.

The high productivity of the construction now appears to have shifted the distribution of the data. In the tokens extracted from the 19th century corpus, which were only those denoting obligation rather than ambiguous possession with responsibility, the four most frequent lexical verb types are now hacer ‘to do’, ser ‘to be’, hablar ‘to speak’, and dar ‘to give’. Hacer remains the single most frequently recurring type, but its presence at the top of the list is perhaps no longer due to the inherent sense of responsibility it contributed in the conventionalized 15th-18th century usage tener que hacer ‘to have (something) to do’. Instead, the high frequency of hacer in [tener que + Infinitive] is potentially due to nothing more than the high frequency and productivity of hacer itself, which also appears in conventionalized expressions. Of the 30 tokens including hacer in the 19th century, 4 contain hacer un viaje ‘to go on a trip’ and one includes hacer el papel ‘to play the role’ as in (126) and (127), respectively.
Siento mucho que no me trajeras la edición de 1527. Tendré que hacer yo mismo un viaje a Madrid.

I am very sorry that you did not bring me the 1527 edition. I will have to make a trip to Madrid myself.

[19th c.; Doña Perfecta: 45]

Si me voy con el estado mayor, tengo que hacer el papel de mirón en una cosa que no entiendo

If I stay with the elders, I must play the part of an onlooker in a matter I do not understand [19th c.; Pepita Jiménez: 84]

Ser ‘to be’ and hablar ‘to speak’, the second and third most frequently recurring lexical verb types respectively, are intransitive. This provides further evidence that a semantic link to the lexical origin of the construction, possession of an object, is no longer a necessary component of the [tener que + Infinitive] construction. The recurrence of ser ‘to be’, which involves no directed action, and instead describes a state, is perhaps the clearest indication of this shift. Examples of [tener que + ser] like that in (128), which is accompanied by the adjective morena ‘dark’ indicate that the construction need not include a transitive infinitive or a possessed object, and instead conveys a generalized sense of obligation.

Morena tiene que ser la tierra para claveles

‘The soil must be dark for carnations’

[19th c.; Davies 2002-]

A final indication of the productivity of the tener que construction is the portion of the total number of tokens that is accounted for by the four most frequent infinitives. As discussed in Section 3.2.6, this percentage decreased from 64% of the data in the 15th to only 26% of the 18th century data. For the 19th century, tokens containing any of the four most frequent lexical
infinitives only account for 16% (76/479) of the occurrences of [tener que + Infinitive]. In fact, these four infinitives, hacer ‘to do’, ser ‘to be’, hablar ‘to speak’, and dar ‘to give’ are the only verbs to appear more than ten times. Figure 4-3 illustrates the almost linear decline in the portion of data accounted for by the four most frequent infinitives over the span of five centuries.

Figure 4-3. Percentage of total data accounted for by the four most frequent lexical infinitives in each century from the 15th to the 19th

4.5. Variationist analysis of the alternating modal variants

The increased productivity of [tener que + Infinitive] and its augmented frequency relative to the three alternative periphrastic modal constructions make it immediately apparent that the tener que construction has made great gains in its distribution throughout the grammar. As it has continued to grammaticalize into the 19th century, [tener que + Infinitive] has evidently encroached on the semantic and functional domain of deontic modality that was previously
dominated by \([haber de + Infinitive]\) and \([deber (de) + Infinitive]\). But what remains to be seen is the potential effect of specific linguistic factors on its development. That is, what quantifiable contextual elements are driving the selection of \([tener que + Infinitive]\) over the alternative constructions as an expression of obligation?

In order to answer this question, a series of multivariate analyses were conducted using the Goldvarb X statistical application (Sankoff, Tagliamonte, and Smith 2005). Such an analysis allows for operationalizing and comparing the effects of a number of different independent variables on the dependent variable, the occurrence of the four modal verb constructions at issue here. Having established the functional overlap of the modal constructions, they can be treated as equivalent variant forms used for the expression of obligation. The data were coded (classified) as belonging to one of a set of factors in several factor groups (independent variables), outlined in the following sections. This methodological approach is aimed at determining the influence of the context in which a given token occurs on the choice of that particular construction (e.g. \([tener que + Infinitive]\)) over the alternative variants.

### 4.6. Coding and factor groups

The combined 2162 tokens of all four modal constructions were coded for a number of linguistic factors with the goal of discerning the relative impact of each on the use of \([tener que + Infinitive]\) in the 19th century. These include: grammatical person, modal verb tense/aspect, polarity, object presence and placement, and document genre/type. Each factor group is treated in turn in the following sections.
4.6.1. Grammatical person

Traugott argues that when grammaticalizing main verbs acquire modal meaning, these meanings commonly become increasingly “focused on the internal world of the speaker's belief and knowledge states.” (1989: 37). Taking the crosslinguistic tendency of evolution from deontic to epistemic modality as an example (cf. Bybee, Perkins and Pagliuca 1994), Traugott posits that expressions of deontic modality are relatively less subjective (i.e. more objective) than expressions of epistemic modality. Furthermore, propositions imposed by external necessity as in (129) are less subjective than those based on (internal) personal conclusions or suppositions (130).

(129)  
*Figúrese usted que recibe el susto, y que además le quedan algunos huesos quebrantados, sin nada de heridas graves, se entiende... pues en tal caso, o se acobarda y huye de Orbajosa, o se tiene que meter en la cama por quince días.*

Imagine that he gets frightened, and in addition he has a few bones broken, without any serious wounds, of course... Well in that case, either his courage will fail him and he will leave Orbajosa, or he will have to keep to his bed for fifteen days. [19th c.; *Doña Perfecta*: 200]

(130)  
*Le había abierto ella misma, sin preguntar quién era, segura de que tenía que ser él.*

“She had opened [the door] herself, without asking who it was, certain that it must be him.” [19th c.; *Regenta*: 541]

Based on the data presented here, it appears that [*tener que* + Infinitive] has only just established itself as a construction consistently and frequently applied to encoding obligation in the interim between the 18th and 19th centuries. Therefore, it is to be expected that the shift from less subjective to more subjective contexts in obligation tokens of the *tener que* construction has not progressed as far as with the older constructions [*deber (de)* + Infinitive] and [*haber de* + Infinitive] that are more commonly associated with being applied to epistemic modality.
Trugott (1989; 1995) asserts that subjectification is integral to grammaticalization, and is evident in numerous paths of semantic change. She argues for three general tendencies of subjectification, of which the first is for meanings to transfer from a basis in the external described situation to a basis in the internal (evaluative/perceptual/cognitive) described situation (Trugott 1989: 34). This tendency is applicable to changes from concrete to abstract meanings, as with the evolution of possession to obligation (Trugott and Dasher 2002: 95). In the present instance, [tener que + Infinitive] is the more recently and actively grammaticalizing construction acquiring an obligation meaning, whereas the two older constructions with haber de and deber (de) have thoroughly acquired modal meaning. They have, in fact, begun to move further into the domain of epistemic modality. Therefore, it is possible that the increased subjectivity of the tener que construction will be manifest in the data.

Aaron and Torres Cacoullos (2005) operationalize subjectivity via a set of measurable variables, including grammatical person. Their results found evidence that “subjectivity is manifested structurally in the tendency to occur with a first-person singular subject (the speaker)” (Aaron and Torres Cacoullos 2005: 22). Presumably, this demonstrates the encoding of speaker involvement (internal evaluation and/or perception) suggested by Traugott’s proposed tendency of semantic change.

To test the hypothesis that the occurrence of the grammaticalizing [tener que + Infinitive] construction will be associated with subjectification, all tokens were coded for their grammatical person and number. Tokens for which the modal verb is non-finite were excluded from this coding, as it was not possible to establish the grammatical subject. Table 4-5 displays the distribution of the data by subject person and number.
Table 4-5. 19th century tokens of the modal variants by grammatical person and number (percentage of total tokens comprising each variant in parentheses)\textsuperscript{34}

<table>
<thead>
<tr>
<th>Person/number</th>
<th>tener que</th>
<th>haber de</th>
<th>deber (de)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>117 (31%)</td>
<td>136 (36%)</td>
<td>130 (34%)</td>
<td>383</td>
</tr>
<tr>
<td>1PL</td>
<td>40 (30%)</td>
<td>43 (32%)</td>
<td>50 (38%)</td>
<td>133</td>
</tr>
<tr>
<td>2SG</td>
<td>31 (16%)</td>
<td>79 (40%)</td>
<td>88 (44%)</td>
<td>198</td>
</tr>
<tr>
<td>2PL</td>
<td>3 (60%)</td>
<td>0</td>
<td>2 (40%)</td>
<td>5</td>
</tr>
<tr>
<td>3SG</td>
<td>208 (21%)</td>
<td>292 (29%)</td>
<td>512 (51%)</td>
<td>1012</td>
</tr>
<tr>
<td>3PL</td>
<td>49 (22%)</td>
<td>73 (33%)</td>
<td>100 (45%)</td>
<td>222</td>
</tr>
<tr>
<td>Total</td>
<td>448 (23%)</td>
<td>623 (32%)</td>
<td>882 (45%)</td>
<td>1953</td>
</tr>
</tbody>
</table>

The tener que construction accounts for 23% of the total data (448 of 1953 tokens), but exhibits a higher rate of occurrence with both first-person singular and first-person plural subjects (31% and 30% respectively). It can be argued that first-person plural subjects are less subjective than first-person singular given that they refer to both the speaker and one or more individuals outside of the speaker (Scheibman 2002). Still, both of the first-person subjects exhibit higher rates of tener que tokens than the data in general, and first-person plural subjects only make up 26% of the total first-person subjects in the data (133/516 tokens); both singular and plural are treated together in the multivariate analyses (Section 4.7). At any rate, the pattern in Table 4-5 alludes to the subjectivity associated with the grammaticalizing tener que construction.

4.6.2. Polarity

In Chapter 3, it was noted that 30% (122/409) of the tokens of [tener que + Infinitive] in the 18th century were negated. Given this high rate of negation in the previous century and its potential facilitation of the bleaching of the possession meaning of tener, polarity may also be an influential factor in the construction’s grammaticalization in the 19th century. However, the data for the 18th century included a large number of tokens of the lexical relative clause construction

\textsuperscript{34} [haber que + Infinitive] is not included in Table 4-5 as the grammatical subject for this construction is invariably third person singular.
and comparatively few tokens that could be interpreted as encoding obligation. In the 19th century, 91% (479/529) of the extracted tokens of [tener que + Infinitive] were obligation occurrences, and when these were coded for polarity, only 5% (26/479) were negated.

This is a departure from the considerably higher rate of negation in the 18th century, but may be in keeping with the proposal by Givón (1978) that negative contexts tend to be a locus of syntactic conservatism and the potential for this to be linked with subjectivity (see Section 1.2.2). Poplack and Dion (2009) provide some evidence to support this; negative contexts were found to strongly favor retention of the conservative synthetic future in their analysis of future temporal reference in French. Torres Cacoullos (2012) also found that negative contexts disfavored the occurrence of the innovative Spanish progressive construction. To test the potential conservatism of negative contexts, all tokens of the four variants were coded as affirmative, negative, or interrogative, as in (131-133) respectively.

(131) ¿Se marchó el carabinero? [...] ¡Es que tengo que pasar contrabando!

Did the customs guard leave? It’s that I have to smuggle contraband!
[19th c.; ¡El siete!: scene 5]

(132) y no había que hacerse ilusiones, ni buscar nueva manera de vivir

‘and he should not entertain any illusions or go in search of a new life’
[19th c.; Regenta: 470]

(133) ¿Cómo he de temer la muerte cuando deseo morir?

‘How am I supposed to fear death when I wish to die?’
[19th c.; Pepita Jiménez: 98]

Interestingly, the rates at which the three other modal constructions appear in negative contexts were higher than that of [tener que + Infinitive], as presented in Table 4-6.
Table 4-6. 19<sup>th</sup> century negation rates for four variant modal constructions

<table>
<thead>
<tr>
<th>Person/number</th>
<th>tener que</th>
<th>haber de</th>
<th>deber (de)</th>
<th>haber que</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>447 (27%)</td>
<td>405 (24%)</td>
<td>724 (43%)</td>
<td>97 (6%)</td>
<td>1673</td>
</tr>
<tr>
<td>Negative</td>
<td>26 (7%)</td>
<td>93 (27%)</td>
<td>156 (45%)</td>
<td>75 (21%)</td>
<td>350</td>
</tr>
<tr>
<td>Interrogative</td>
<td>6 (4%)</td>
<td>125 (90%)</td>
<td>8 (6%)</td>
<td>0</td>
<td>139</td>
</tr>
<tr>
<td>Total</td>
<td>479 (22%)</td>
<td>623 (29%)</td>
<td>888 (41%)</td>
<td>172 (8%)</td>
<td>2162</td>
</tr>
</tbody>
</table>

The *tener que* construction exhibits the lowest rate of appearance in negative contexts, as it is three times as likely to appear in the overall data as it is in the negated tokens. Perhaps surprisingly, *haber que* appears more frequently in negative contexts than does *tener que* (accounting for 21% of the negated tokens), despite the fact that it is arguably just as innovative a form as [*tener que + Infinitive*]. Still, *haber que* is represented by the fewest tokens and is limited in its productivity; [*haber que + Infinitive*] appears only with third person singular subjects and as mentioned in 4.1.1, this particular construction did not necessarily evolve directly from possession to obligation as with [*tener que + Infinitive*]. With this in mind, it is perhaps best not to discard the hypothesized conservatism associated with negation and instead view the *haber que* pattern as potentially anomalous.

4.6.3. Tense/aspect

According to Hopper (1982), one of the principal functions of aspect is to distinguish events from non-events. That is, it serves to order and foreground distinct events in a sequence from non-event background information that is not part of that sequence. Consider, for example, the use of aspect in (134).

(134)  *La emoción la ahogaba, y tuvo que sentarse sobre una piedra.*

‘The feeling was suffocating her, and she had to sit down on a rock.’ [19<sup>th</sup> c.; Davies 2002-]
The Imperfect (past imperfective) in *la ahogaba* ‘was suffocating her’ provides background information about an ongoing emotional condition while the perfective Preterit in *tuvo que sentarse* ‘she had to sit down’ explains the event that punctuates that ongoing state. This use of the perfective aspect then serves to present the subject’s sitting down as a bounded, and consequently foregrounded, event (Bybee et al. 1994: 90, 301).

There is also evidence of an association between this foregrounding function of perfective aspect and the appearance of an innovative grammaticalizing form. Klein-Andreu (1991) found that the historically innovative periphrastic pluperfect construction in Spanish was preferred for foregrounding. Furthermore, Aaron and Torres Cacoullos (2005) propose an explanation of the link between past events and speaker perspective based on the distinction between realized and non-realized situations. In their case, non-realized (irrealis) events disfavored the occurrence of a middle-marked verb form (*salirse* ‘to go out’). They argue that the tendency of the form to appear in past tense contexts was a manifestation of subjectivity: past events have already been realized are therefore less associated with presupposition (2005: 21-22). Additionally, they found that negative polarity disfavored the occurrence of the middle-marked form, which provides another link to the arguments about conservatism in negative contexts discussed above in 4.6.2.

Given this suggestive evidence, to test the hypothesis that perfective aspect will correlate with the appearance of *[tener que + Infinitive]*, the data were coded for their tense and aspect. The numerous individual tense categories were then conflated due to the distribution of the data among these tokens. As Table 4-7 illustrates, the vast majority of the tokens were confined to three tense/aspect categories: Preterit (past perfective), Imperfect (past imperfective), and Present (imperfective present). Tokens not categorized as Present, Imperfect, or Preterit are grouped together as “other” given that aside from the three aforementioned categories, no other individual
tense/aspect category (e.g. Future, Present Perfect, etc.) accounted for more than 5% of the data for a given construction.

In Table 4-7, it is immediately apparent that Present tokens are by far the most frequent, accounting for more than half of the data (1261/2162 tokens). However, the innovative [tener que + Infinitive] is less likely to appear in Present and Imperfect contexts than it is overall (15% and 18% vs 22%, respectively). The specific tense that deviates most sharply from the general distribution is the perfective Preterit, in which tener que accounts for 50% of the tokens\textsuperscript{35}. As it did with polarity, [haber que + Infinitive] appears to pattern more closely with the two older constructions than with [tener que + Infinitive], even though it is also innovative: haber que accounts for a smaller percentage of the perfective tokens than of the overall data.

\textbf{4.6.4. Object presence and placement}

In Chapter 3, the importance of a direct object to determining the meaning of a given occurrence of the [tener que + Infinitive] construction was discussed. In the intermediate

\textsuperscript{35} The category “other” in Table 4-7 includes a number of imperfective tenses, including the Present Perfect, in which the haber de and haber que constructions are unlikely to appear. Although it is striking that [tener que + Infinitive] accounts for half of these tokens, it is not as appropriate to consider this a specific tense, as it is with the Preterit (past perfective).
semantic stage considered there, a direct object intervening between tener and the relative pronoun que was taken as an indication that such a token of the construction encoded possession with a potential implication of responsibility or intention, as in (135).

(135) Estoy pensando en que cuando uno tiene ALGO que guardar, las llaves le parecen una falsa garantía.

‘I am thinking about how when one has SOMETHING to protect, keys seem a false security.’
[19th c.; Las circunstancias: Scene 1]

This placement contrasts with the placement of an object outside of the construction, either prior to tener or following the lexical infinitive ((136) and (137), respectively). In cases for which the lexical infinitive is intransitive, there is no direct object at all (138).

(136) UN FAVOR tenía que pedirte

‘I had to ask you A FAVOR’ [19th c.; Amor de padre: Scene 7]

(137) todos sabemos que tenemos que llevar NUESTRA CRUZ sobre los hombros

‘we all know that we must carry OUR CROSS on our shoulders’
[19th c.; Davies 2002-]

(138) Eran hombres ocupados que tenían que madrugar.

‘They were busy men who had to rise early.’ [19th c.; Regenta: 262]

In all three examples, the interpretation of obligation is unambiguous. Given that the modal construction evolved from a possession construction which necessarily included a direct object, tokens including an intransitive infinitive are assumed to be innovative. With that in mind, the data were coded as exhibiting a preposed object, a postposed object, or none at all (the three
arrangements in ((136-138), respectively). Assuming that the older *haber de* and *deber (de)* constructions have been established with intransitive lexical infinitives since long before the 19th century, while the inclusion of such infinitives is presumably more recent in *[tener que + Infinitive]*, tokens with explicit direct objects are expected to be more favorable to the occurrence of the *tener que* construction.

### 4.6.5. Source/genre

As mentioned in Section 4.3, all of the 19th century data were extracted from written prose. However, there is some variability in the types of sources from which tokens were drawn. The majority of the compiled corpus consists of novels, but also includes a number of theatrical works and the 19th century portion of the *Documentos Lingüísticos de la Nueva España* (Company Company 1994), written in the form of letters reporting magisterial and judicial proceedings in Mexico. The majority of the supplemental data from Davies (2002-) online *Corpus del Español* were also taken from novels, but includes two non-fiction works: *Compendio de la vida de San Alfonso María de Ligorio* is the life story of a Catholic bishop, and *Manual del viajero y guía de los forasteros en Valencia* is a traveler’s guide to Valencia, Spain.

As all of the texts are written in prose, the major distinction in terms of genre is between the theatrical plays and the literary works. There is reason to believe that the writing style of the plays may be distinct from that of the novels, given the potential effort on the part of an author to emulate a particular vernacular and/or portray certain characters as belonging to one social class or another. Indeed, Francisco Martínez de la Rosa, the author of *Amor de padre* (one of the plays included in the corpus), indicates in a foreword to the play that he seeks to avoid affectation and strives to achieve simplicity in his writing.
There is evidence suggesting that theater may in fact approximate spoken language more closely than other written language. In their analysis of Brazilian Portuguese future expression, Poplack and Malvar (2007) found that their collection of 20th century popular plays patterned closely with 20th century speech in the distribution of several variant constructions expressing future temporal reference. In light of this result, it is to be expected that data drawn from theater will be less conservative (i.e. favor the use of [tener que + Infinitive]) than that of the other sources. To test this hypothesis, tokens were coded for their source: novels (including nonfiction works from Davies’ corpus), plays, and Documentos Lingüísticos de la Nueva España.

4.6.6. Exclusions

Given the distribution of the data (presented in tables 4-5, 4-6, and 4-7 above), a number of factors were either excluded from the analysis or combined with others. This was done for two reasons. First, tokens were scarce for several factors (e.g. there were only 5 total occurrences of a 2nd person plural subject) and combining them into larger groups provided higher token counts yielding more reliable statistical measures. Second, simplifying the factor groups into binary pairs allows for a direct comparison of the range of factor weights for each group. The difference between the highest and lowest factor weights can be interpreted as the magnitude of the effect of that group of factors (relative to the other groups considered in the multivariate analysis).

For the factor group Subject, singular and plural 1st person subjects were combined into a single factor; tokens with all other subjects/numbers were grouped into a second. As noted in Section 4.6.3, tokens of the Imperfect, Present, and Preterit made up 90% of the data; other tenses were therefore excluded. Polarity was also simplified to two factors, affirmative and negative, with interrogatives excluded. Interrogative tokens made up only 1% or less of the data for all variants except for [haber de + Infinitive]. For the factor group Source, only 16 tokens were
drawn from the Documentos Lingüísticos de la Nueva España, so these, the data from novels, and all of the supplemental data from Davies’ online corpus were counted as Literature. The remaining tokens were from plays. Finally, the Object factor group was divided into tokens for which a direct object was present and tokens for which it was absent.

4.7. Results

Multivariate analyses were conducted in order to determine which factors influenced the choice of the tener que construction over the three alternative variants in the expression of obligation. In three individual analyses, [tener que + Infinitive] was paired with one of the other forms. A fourth analysis included all four constructions, treating tener que as the application value while combining the other three forms and treating tokens of those constructions as non-applications of the variable rule.

4.7.1. Individual analyses

The results of the multivariate analyses pairing the [tener que + Infinitive] with deber (de), haber de, and haber que respectively are presented below in Table 4-8. Each analysis is presented in its own column, facilitating direct comparison of the effect from each factor group. Factor weights range from 0 to 1, with weights above 0.5 favoring the appearance of [tener que + Infinitive]; factor weights below 0.5 disfavor its use. The range represents the relative magnitude of effect for each factor group, with a larger range indicating that a given factor group is relatively more influential in accounting for the variation than a factor group with a smaller range.
Table 4-8. Results of multivariate analyses of factors contributing to choice of \([tener que + \text{Infinitive}]\) as opposed to each of the alternative variants in the 19th century. Factor weights indicate the contribution of factors selected as significant; brackets indicate that a factor was not selected as significant in a given analysis.\(^{36}\)

<table>
<thead>
<tr>
<th>Factor group</th>
<th>vs. deber (de)</th>
<th>vs. haber de</th>
<th>vs. haber que</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 1160 (Input: 0.29)</td>
<td>N = 835 (Input: 0.42)</td>
<td>N = 525 (Input: 0.72)</td>
<td></td>
</tr>
<tr>
<td><strong>Polarity</strong></td>
<td><strong>Factor weight</strong></td>
<td><strong>Factor weight</strong></td>
<td><strong>Factor weight</strong></td>
</tr>
<tr>
<td>Affirmative</td>
<td>0.56</td>
<td>0.81</td>
<td>0.62</td>
</tr>
<tr>
<td>Negative</td>
<td>0.21</td>
<td>0.45</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Range:</strong> 35</td>
<td><strong>Range:</strong> 36</td>
<td><strong>Range:</strong> 53</td>
<td></td>
</tr>
<tr>
<td><strong>Aspect</strong></td>
<td><strong>Factor weight</strong></td>
<td><strong>Factor weight</strong></td>
<td><strong>Factor weight</strong></td>
</tr>
<tr>
<td>Perfective</td>
<td>0.78</td>
<td>0.55</td>
<td>0.76</td>
</tr>
<tr>
<td>Imperfective</td>
<td>0.46</td>
<td>0.22</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Range:</strong> 32</td>
<td><strong>Range:</strong> 33</td>
<td><strong>Range:</strong> 32</td>
<td></td>
</tr>
<tr>
<td><strong>Person</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>0.64</td>
<td>0.57</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>0.45</td>
<td>0.47</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Range:</strong> 19</td>
<td><strong>Range:</strong> 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Object presence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>[ ]</td>
<td>[ ]</td>
<td>0.61</td>
</tr>
<tr>
<td>Present</td>
<td>[ ]</td>
<td>[ ]</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Range:</strong> 20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor not selected as significant: Source

What is perhaps most interesting is that polarity is the most influential factor group for all three analyses. Affirmative contexts favor the use of \([tener que + \text{Infinitive}]\), while negative contexts disfavor it. This lends credence to the notion that negative contexts tend to be more conservative, preserving the older constructions with \(deber (de)\) and \(haber de\). Surprisingly, polarity appears to have the strongest effect in the choice between \([tener que + \text{Infinitive}]\) and \([haber que + \text{Infinitive}]\), as negative contexts strongly disfavor the former construction. The latter is an innovative form, so the same explanation used for the \(deber (de)\) and \(haber de\) constructions is not applicable.

---

\(^{36}\) Person was not included as a factor group in the analysis of \([haber que + \text{Infinitive}]\), as it appears invariably with a 3SG subject.
However, this piece of evidence may help to clarify the issue of the genesis of [haber que + Infinitive]. While Yllera (1980) suggests that its origin is analogical to that of [tener que + Infinitive], the differences in the distributions (see Sections 4.3 and 4.6) of the two constructions point to a different explanation. In the case of [haber que + Infinitive], some of the earliest tokens already exhibit *hay* ‘there is’ rather than the lexical *haber* ‘to have’, and as mentioned, the subject is invariably third person singular in the 19th century data. It seems that the [haber que + Infinitive] construction did not evolve via the same semantic pathway as [tener que + Infinitive].

A second consistent finding is that Aspect is the factor group found to be the second most influential in selecting [tener que + Infinitive] over each of the three alternative constructions. That is, perfective contexts favor the appearance of [tener que + Infinitive] in each case. In fact, the range between the factor weights in this group (signaling the magnitude of its effect) is nearly equal to that of Polarity in all but the analysis with [haber que + Infinitive]. This also supports the hypothesis that the foregrounding effect of perfective contexts would favor the innovative *tener que* construction.

Grammatical person was selected as the third significant factor group in the analyses with [deber (de) + Infinitive] and [haber de + Infinitive], the two older variants. The presupposed favoring effect of first person subjects is evident, albeit with a smaller magnitude of effect (ranges of 19 and 10, respectively). This is an additional area in which the *haber que* construction patterns differently from [tener que + Infinitive]; the invariable third person singular subject of *haber que* meant that this factor group was excluded from that analysis.

A final piece of evidence regarding the distinct distributions of [tener que + Infinitive] and [haber que + Infinitive] is found in the selection of object presence as a significant factor group in the analysis of these two constructions. Though not significant in the other two analyses, explicit direct objects disfavored the appearance of the *tener que* construction, suggesting that this is another area in which *haber que* is more limited in its distribution; *tener que* is evidently more
likely to appear with intransitive infinitives. This is potentially a product of the association of the 
haber que construction with the notion of existence, e.g. nada hay que temer ‘there is nothing to 
fear’, as in example (113) above (see Section 4.1.1). In such an instance, there is necessarily an 
object of the transitive temer ‘to fear’: that entity which is said to exist (or not exist).

4.7.2. [tener que + Infinitive] vs all three alternative variants

Polarity and Aspect were both found to be significant contributing factors, and were 
ranked as the two most influential independent variables, respectively, for each of the three paired 
analyses. Given this consistency, it appears appropriate to treat the variation characterizing the 
evolution of [tener que + Infinitive] as that of an innovative construction belonging to a general 
system of variant constructions for the expression of obligation. Therefore, a final multivariate 
analysis was conducted, grouping the three alternative forms together. This was done to make a 
final assessment of the system as a whole and the factors contributing to the emergence of [tener 
que + Infinitive] as an increasingly frequent means of encoding deontic modality. Table 4-9 
displays the results of this analysis.
Table 4-9. Results of multivariate analysis of factors contributing to the selection of [tener que + Infinitive] over [deber (de)/haber de/haber que + Infinitive]

Total N = 1800, Corrected mean: 0.17 (20% tener que)

<table>
<thead>
<tr>
<th>Factor weight</th>
<th>% tener que</th>
<th>Total N</th>
<th>% data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>0.58</td>
<td>23%</td>
<td>1470</td>
</tr>
<tr>
<td>Negative</td>
<td>0.20</td>
<td>5%</td>
<td>330</td>
</tr>
<tr>
<td><strong>Range:</strong></td>
<td></td>
<td><strong>38</strong></td>
<td></td>
</tr>
<tr>
<td>Perfective</td>
<td>0.79</td>
<td>50%</td>
<td>160</td>
</tr>
<tr>
<td>Imperfective</td>
<td>0.47</td>
<td>17%</td>
<td>1640</td>
</tr>
<tr>
<td><strong>Range:</strong></td>
<td></td>
<td><strong>32</strong></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>0.64</td>
<td>29%</td>
<td>415</td>
</tr>
<tr>
<td>Other</td>
<td>0.46</td>
<td>17%</td>
<td>1385</td>
</tr>
<tr>
<td><strong>Range:</strong></td>
<td></td>
<td><strong>18</strong></td>
<td></td>
</tr>
</tbody>
</table>

Factors not selected as significant: Object presence, Source

The results presented in Table 4-9 reveal no major differences in the direction or magnitude of effect of the factor groups from the individual analyses above. As before, affirmative contexts favor the appearance of [tener que + Infinitive]. It is worth noting, however, that the relatively large range found for this factor group is due to the fact that negative contexts strongly disfavor its appearance. This is in contrast with the results found for Aspect, which again show that perfective aspect strongly favors tener que, while imperfective aspect only mildly disfavors it. Person is once again selected as significant, but has the smallest effect. These results confirm that the generalizations drawn from the individual analyses hold true when analyzing the system as a whole.
4.8. Conclusion

The tendencies present in the 19th century data are in stark contrast to the patterns found in previous centuries. Generally speaking, the variables that appeared to be influencing the evolution of [tener que + Infinitive] in the 18th century and earlier are no longer as potent as they were. In the 15th-18th centuries, it was evident that the lexical infinitives with which the construction appeared were providing an important semantic component, arguably contributing to pragmatic inference and subsequent meaning change. In the 19th century, recurring individual lexical verb types are no longer as frequent as they were, accounting for an even smaller portion of the data than they did in the 18th, where they were already declining.

What is more, the source of the data (literature vs. theater) was not found to be statistically significant in accounting for the variation among the four modal verb constructions while the variables that were significant were three grammatical factor groups, Polarity, Aspect, and Person. In another departure from older distributional patterns, negative polarity now disfavors the appearance of [tener que + Infinitive], whereas there were a substantial number of negative tokens of the construction in the 18th century when it still had uses other than expressing obligation. This result supports the hypothesis that negative contexts tend to be conservative with respect to grammaticalizing constructions. The presence of a direct object, a structural/semantic variable, only has a small effect in accounting for the specific variation between the tener que and haber que constructions.

Interestingly, [haber que + Infinitive] does not mirror [tener que + Infinitive] in its frequency or distribution. It is therefore difficult to argue that its genesis is analogous to that of the tener que construction, despite their superficial structural similarities. Although haber and tener share a common semantic source in possession, [haber que + Infinitive] does not appear to have evolved through the same stages as [tener que + Infinitive]. Rather, it seemingly derives its
sense of obligation from the notion of existence, as its modern form exhibits the invariable *hay* ‘there is, there exists’. This is seemingly not unlike the development of a copula into a marker of deontic modality, attested in several languages (Denning 1987; Heine and Kuteva 2002: 97).

Thus, the variation within the obligation system is largely accounted for by grammatical variables, making it appear that *[tener que + Infinitive]* has by this point firmly established itself as a modal construction, participating in the system and varying with the alternative forms in this functional domain. Factors such as the lexical infinitive and the presence of a direct object, variables that evidently played roles in the semantic transition from possession to obligation, appear to no longer be operable. It is as if the construction has grammaticalized to such a degree that its appearance no longer relies on the mechanisms by which it passed through the beginning and intermediate stages of its development (see Chapters 2 and 3, respectively).

Still, it is important to note that this “degree” of grammaticalization is best viewed as a continuous variable, and is relative to earlier, overlapping forms and meanings of the current manifestation. That is, it is not appropriate to say that a construction has completed the process of grammaticalization; *[tener que + Infinitive]* has evidently gone on to develop an epistemic usage in the modern language, and could potentially continue on that trajectory. As another measure of its relative grammaticalization, the *tener que* construction lacks the sort of phonetic reduction exhibited by *have to* (e.g. *have to* > [hæftə]), and even the *have to* construction has been called a “quasimodal” rather than a canonical modal (Brinton 1991; Fischer 1994).

Given that the absolute and relative frequencies of *[tener que + Infinitive]* are clearly on the rise in the transition from the 18th to the 19th century and its meaning is evidently still in flux, it is necessary to analyze 20th century data to determine whether or not these trends are continued in the modern language. The following chapter will do just that, allowing for a comparison of the

---

37 The construction does at times appear in the Imperfect and the Preterit as *había que* and *hubo que* ‘it was necessary to…’, but is still invariable in its person and number.
factors contributing to the appearance of [tener que + Infinitive] in the 20th century to the results reported here for the 19th. Specifically, the hypothesis of continued subjectification as [tener que + Infinitive] grammaticalizes will be tested (per Traugott 1989). The 19th century data suggest that this may be true (1st person subjects favored the appearance of [tener que + Infinitive]), but a confirmation of the hypothesis will require comparison with the 20th century. Subsequent analyses will seek this and other signs of changes in the variation among these four modal constructions with the goal lending further quantitative support to the testable hypotheses offered by previous treatments of crosslinguistic tendencies in grammaticalization.
Chapter 5

Variation between [tener que + Infinitive] and other periphrastic modal constructions: the 20\textsuperscript{th} century

In the previous chapter, the [tener que + Infinitive] construction was analyzed as one of several periphrastic modal verb constructions, each variably applied to the expression of obligation. The transition from the 18\textsuperscript{th} to the 19\textsuperscript{th} century was marked by substantial increases in both the overall token frequency of the tener que construction and its frequency relative to those of the other constructions in the obligation system as a whole. This was taken as evidence that [tener que + Infinitive] had progressed far enough along its cline of grammaticalization to thoroughly establish itself as a modal construction.

This is not to say, however, that the construction is assumed to have completed the process of becoming a modal verb. On the contrary, the [tener que + Infinitive] construction continues to undergo developments in its distribution and frequency in the interval from the 19\textsuperscript{th} century to the 20\textsuperscript{th} century. The present chapter focuses on the evident changes that have resulted in the construction’s distribution in the modern language, utilizing 20\textsuperscript{th} century sources. Data were extracted from a combination of written texts and transcribed interviews. These were analyzed in the same manner as the 19\textsuperscript{th} century data (see Chapter 4), revealing that the tener que construction has once again gained frequency. What is more, this increase is at the expense of the modal constructions that predate tener que ([deber (de) + Infinitive] and [haber de + Infinitive]).

The chapter is arranged as follows: first, the frequency and distribution of the [tener que + Infinitive] construction itself are discussed. This is followed by an assessment of the relative frequency of the target construction with respect to the alternative modal constructions used to encode obligation. The results of multivariate analyses of the variation within the obligation
system are then presented, allowing for a direct comparison to the linguistic factors that were found to be influencing the appearance of \([tener que + Infinitive]\) in the 19th century.

5.1. 20th century corpus and data

As mentioned in Chapter 4, \([tener que + Infinitive]\) underwent a substantial increase in its absolute frequency as a modal verb construction between the 18th and 19th centuries. This was mirrored by a dramatic decrease in the rate at which the older construction \([haber de + Infinitive]\) was applied to expressing obligation. Given the apparent trajectory of the tener que construction’s expansion, it is to be expected that the 20th century will provide further evidence of a consistent increase in frequency.

To compare the previous evidence with data from modern Spanish, a corpus was once again compiled, but with one crucial difference: transcriptions of oral interviews were included in addition to written texts. The advantage of utilizing spoken language in the corpus is that it allows for a comparison of two distinct linguistic mediums. Presuming that written language is likely to be more conservative (and adherent to prescriptive norms) than spoken language, it is expected that the transcribed interviews will yield a different distribution (in terms of overall rates, not necessarily in terms of the effects of conditioning factors) of the four modal constructions than that of written texts.

Another difference in the 20th century corpus is that two distinct varieties of Spanish are represented. In Chapters 3 and 4, the Documentos Lingüísticos de la Nueva España (Company Company 1994) were included as a source of data. These documents exemplify Spanish from what is now Mexico from the 16th to the 19th century and yield tokens of the \([tener que + Infinitive]\) that illustrate the semantic and distributional changes discussed in the previous two chapters. Still, the source of the data (documents from Mexico, theatrical plays, or novels) was
not found to be a significant factor influencing the variation in the obligation system for the 19th century.

Despite the lack of corroborating evidence from the 19th century data presented here, the data sources available for 20th century make it possible to analyze a much larger representative sample of Mexican Spanish. This variety was therefore included in the corpus, as reference grammars suggest that the choice of different modal constructions varies from one variety of Spanish to another (Butt and Benjamin 2004: 327-330). Furthermore, Fairclough (2000) found quantitative evidence suggesting an ongoing change in the distribution of alternative modal variants in Mexican Spanish.38

In all, three sets of interviews were included in the corpus: *El habla de la ciudad de México* (Lope Blanch 1971) and *El habla popular de la ciudad de México* (Lope Blanch 1976) represent educated and vernacular speech from Mexico city, respectively, and *El habla de la ciudad de Madrid* (Esgueva and Cantarero 1981) represents educated speech from Madrid, Spain. The written data are drawn from eight literary works (four from each variety) all written in prose, by a total of seven different authors. The texts include two books (one a collection of short stories and the other a novel) by the Mexican author Juan Rulfo, while the remainder were drawn from Davies’ (2002-) online *Corpus del Español*. Table 5-1 displays the makeup of the corpus.

38 Fairclough compared Mexican Spanish speakers with Mexican-American Spanish speakers in Houston, Texas.
Table 5-1. Corpus of 20th century texts and interviews (authors for written works in parentheses)

<table>
<thead>
<tr>
<th>Written texts (Mexico)</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Gazapo] Gazapo (Gustavo Sainz)</td>
<td>42,181</td>
</tr>
<tr>
<td>[Artemio] La muerte de Artemio Cruz (Carlos Fuentes)</td>
<td>94,648</td>
</tr>
<tr>
<td>[Pedro] Pedro Páramo (Juan Rulfo)</td>
<td>33,400</td>
</tr>
<tr>
<td>[Llano] El llano en llamas (Juan Rulfo)</td>
<td>42,400</td>
</tr>
<tr>
<td><strong>Mexico Written total</strong></td>
<td><strong>212,629</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Written texts (Spain)</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Camino] El camino (Miguel Delibes)</td>
<td>52,516</td>
</tr>
<tr>
<td>[Julián] La reindivición del conde Don Julián (Juan Goytisolo)</td>
<td>51,792</td>
</tr>
<tr>
<td>[Silencio] Tiempo de silencio (Martín Santos)</td>
<td>100,971</td>
</tr>
<tr>
<td>[Kensington] Kensington Gardens (Xavier B. Fernández)</td>
<td>19,963</td>
</tr>
<tr>
<td><strong>Spain Written total</strong></td>
<td><strong>225,242</strong></td>
</tr>
<tr>
<td><strong>Written total</strong></td>
<td><strong>437,871</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interviews (Mexico)</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>[MexCult] El habla de la ciudad de México (Lope Blanch 1971)</td>
<td>175,000</td>
</tr>
<tr>
<td>[MexPop] El habla popular de la ciudad de México (Lope Blanch 1976)</td>
<td>177,000</td>
</tr>
<tr>
<td><strong>Interviews (Spain)</strong></td>
<td><strong>144,000</strong></td>
</tr>
<tr>
<td>[MadCult] El habla de la ciudad de Madrid (Esgueva and Cantarero 1981)</td>
<td><strong>496,000</strong></td>
</tr>
<tr>
<td><strong>Spoken total</strong></td>
<td><strong>496,000</strong></td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td><strong>933,871</strong></td>
</tr>
</tbody>
</table>

5.2. Distribution of data in the corpus

All tokens of the four periphrastic modal constructions (tener que, haber que, haber de, and deber (de)) were extracted from the corpus. Those tokens that expressed obligation, as in (139) were the focus of the present study, and those that encoded possession (140) or epistemic modality (141) were excluded from the analysis. The excluded tokens did, however, play a role in the interpretation of the data, as discussed below in 5.3.3.
(139) *para tener las cosas, para adquirirlas, hay que luchar para* for have.INF ART things for acquire.INF:them have_to fight.INF for tenerlas
have.INF:OBJ
‘to have things, to acquire them, one has to fight to have them’ 

[20th c.; MexCult: 404]

(140) *Pero no tienen qué darnos de comer.* But NEG have.PRS.3PL anything give.INF:us REL eat.INF
‘But they do not have anything to give us to eat.’ [20th c.; Llano: 93]

(141) *algún ruido debió de hacer porque desperté* some noise must.PST.3SG make.INF because wake_up.PST.1SG
‘it must have made some noise because I woke up’ [20th c.; Davies 2002-

With the exclusion of examples such as those in (140-141), a total of 1992 combined tokens of the four modal constructions were collected from the 20th century corpus. Table 5-2 presents the distribution of the data.

<table>
<thead>
<tr>
<th></th>
<th>Oral</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[MexPop]</td>
<td>[MexCult]</td>
</tr>
<tr>
<td>deber</td>
<td>54</td>
<td>145</td>
</tr>
<tr>
<td>haber de</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>haber que</td>
<td>66</td>
<td>50</td>
</tr>
<tr>
<td>tener que</td>
<td>245</td>
<td>240</td>
</tr>
<tr>
<td>Total N</td>
<td>379</td>
<td>437</td>
</tr>
<tr>
<td># words</td>
<td>177000</td>
<td>175000</td>
</tr>
</tbody>
</table>
Immediately noticeable in the table is the fact that tokens of \([\text{tener que} + \text{Infinitive}]\) are by far the most frequent of any of the four constructions. They make up more than half (1047 of 1992) of the total data and are the most frequent in all of the sources except for the written texts from Mexico. This is a marked increase from the portion of the data made up of \text{tener que} tokens in the 19th century, as illustrated in Figure 5-1.

Comparing the 20th century data with those of the 17th – 19th centuries highlights the consistently increasing relative frequency of the \text{tener que} construction. Starting from a negligible portion of the data in the 17th century, \([\text{tener que} + \text{Infinitive}]\) first made up a substantive portion of the obligation system in the 18th century. In the 19th century, it clearly cemented itself as a modal construction (see Chapter 4). But the 20th century distribution exhibits an even greater increase; the \text{tener que} construction had evidently not yet completed its development, given the concomitant escalation in frequency associated with advanced grammaticalization (Bybee 2010: 3%)

![Figure 5-1. Relative frequencies of modal constructions from the 17th to the 20th Century](image-url)
The amplified rate of appearance of the tener que construction has been mirrored by [haber que + Infinitive], which has again gained ground in the modern data, albeit to a lesser extent; both forms more than doubled their frequencies relative to the other two modal constructions, [haber de + Infinitive] and [deber (de) + Infinitive].

As tener que has increased in frequency, it has done so at the expense of its antecedents in the obligation system. Together, [haber de + Infinitive] and [deber (de) + Infinitive] accounted for the vast majority of the data up to and including the 19th century, but haber de made up a smaller and smaller portion with each passing century. In the 20th century, it has all but disappeared as an expression of obligation, occurring only 68 times in the corpus (3% of the data). The deber (de) construction was the most frequent of the four constructions in the 19th century, but it too has receded, making up little more than a quarter of the 20th century data.

The [tener que + Infinitive] construction makes up a substantially larger portion of the data in the 20th century than in the 19th century, and it has also undergone an increase in its overall token frequency. Table 5-3 depicts this increase, again comparing the rates at which the four modal constructions occur in the 20th century with those of the previous two centuries.

Table 5-3. Frequencies (occurrences per million words) of the four alternative modal constructions in the 18th-20th centuries

<table>
<thead>
<tr>
<th>Century</th>
<th>18th N</th>
<th>Frequency</th>
<th>19th N</th>
<th>Frequency</th>
<th>20th N</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>[deber (de) + Inf.]</td>
<td>45</td>
<td>417</td>
<td>888</td>
<td>792</td>
<td>547</td>
<td>586</td>
</tr>
<tr>
<td>[haber de + Inf.]</td>
<td>111</td>
<td>1028</td>
<td>623</td>
<td>556</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>[haber que + Inf.]</td>
<td>11</td>
<td>102</td>
<td>172</td>
<td>153</td>
<td>329</td>
<td>352</td>
</tr>
<tr>
<td>[tener que + Inf.]</td>
<td>15</td>
<td>139</td>
<td>479</td>
<td>427</td>
<td>1047</td>
<td>1122</td>
</tr>
<tr>
<td>Total N</td>
<td>182</td>
<td></td>
<td>2162</td>
<td></td>
<td>1992</td>
<td></td>
</tr>
<tr>
<td>Total # words</td>
<td>108,000</td>
<td></td>
<td>1,121,221</td>
<td></td>
<td>933,871</td>
<td></td>
</tr>
</tbody>
</table>

The highlighted cells in Table 5-3 point out the rapidly increasing rate at which \textit{tener que} + Infinitive appears in the corpora compiled for the final three centuries analyzed here. The innovative \textit{tener que} construction is nearly ten times as frequent as an expression of obligation in the 20\textsuperscript{th} century as it was in the 18\textsuperscript{th}. This is in stark contrast with the older \textit{haber de} + Infinitive which displays the opposite trend.

\textbf{5.2.1. Written vs. spoken data}

While the corpora utilized for the previous analyses comprised only written documents, the 20\textsuperscript{th} century corpus affords the advantage of comparing spoken data with written data. When viewed side by side, the absolute frequencies of the four modal constructions confirm the hypothesis that the data drawn from spoken language are distributed differently than those from written language.

Table 5-4. Frequencies (per million words) of the four modal variants in the 20\textsuperscript{th} century: written language vs. spoken language

<table>
<thead>
<tr>
<th>Medium</th>
<th>Written</th>
<th>Spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Frequency</td>
</tr>
<tr>
<td>\textit{deber (de) + Inf.}</td>
<td>300</td>
<td>685</td>
</tr>
<tr>
<td>\textit{haber de + Inf.}</td>
<td>50</td>
<td>114</td>
</tr>
<tr>
<td>\textit{haber que + Inf.}</td>
<td>121</td>
<td>276</td>
</tr>
<tr>
<td>\textit{tener que + Inf.}</td>
<td>304</td>
<td>694</td>
</tr>
<tr>
<td>Total N</td>
<td>775</td>
<td></td>
</tr>
<tr>
<td># words</td>
<td>437,871</td>
<td></td>
</tr>
</tbody>
</table>

The two older modal constructions \textit{deber (de) + Infinitive} and \textit{haber de + Infinitive} are both more frequent in the written portion of the corpus, while the two innovative forms \textit{tener que + Infinitive} and \textit{haber que + Infinitive} are both more frequent in the spoken portion. A $\chi^2$
test confirms that the difference between the two distributions is statistically significant ($\chi^2 = 268.2, p < 0.001$). Thus, the written data evidently follows a more conservative pattern than the transcribed interviews, as was expected. When the distinction between the written and spoken data is added to the graph of the relative frequencies presented above in Figure 5-1, the advancement of [tener que + Infinitive] is highlighted even more.

Figure 5-2. Relative frequencies of modal constructions from the 18th to the 20th century, distinguishing written data from spoken data

The data drawn from the transcribed interviews exhibit an even greater advancement of the tener que construction than the written data from the same century. The distribution of the 20th century written data can be viewed as the next stage in the progression suggested by the 18th and 19th century data. Even so, this is evidently a more conservative (older) distribution of the four constructions, than that found in the oral data, where the innovative [tener que + Infinitive] is three times as frequent as [deber (de) + Infinitive], the next most frequent construction.
Furthermore, the haber de construction is effectively nonexistent in the oral data, occurring only 18 times.

5.2.2. Iberian vs. Mexican data

In addition to allowing for the comparison of the spoken and written mediums, the 20th century data represent two varieties of Spanish\(^40\): Mexican and Iberian. Despite not being a significant factor in accounting for the variation among the four modal constructions in the 19th century, it appears that the source of the data plays a more important role in the 20th. When the data are divided based on the region from which they are drawn, they display two distinct patterns, as depicted in Table 5-5.

Table 5-5. Frequencies (per million words) of the modal variants in the 20th century, arranged by variety

<table>
<thead>
<tr>
<th>Variety</th>
<th>Iberian</th>
<th>Mexican</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Frequency</td>
</tr>
<tr>
<td>[deber (de) + Inf.]</td>
<td>161</td>
<td>435</td>
</tr>
<tr>
<td>[haber de + Inf.]</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>[haber que + Inf.]</td>
<td>161</td>
<td>436</td>
</tr>
<tr>
<td>[tener que + Inf.]</td>
<td>392</td>
<td>1068</td>
</tr>
<tr>
<td><strong>Total N</strong></td>
<td>752</td>
<td></td>
</tr>
<tr>
<td><strong># words</strong></td>
<td>369242</td>
<td></td>
</tr>
</tbody>
</table>

The Mexican data illustrate substantially more tokens of [deber (de) + Infinitive] and fewer tokens of the haber de and haber que constructions. The difference between the two

\(^40\) The term ‘variety’ is used here in a general sense. The transcribed interviews represent the speech of two specific cities: Madrid, Spain and Mexico City, Mexico. However, the written portion of the corpus includes authors who were not from these cities. Therefore, these are not specific dialects per se, but are assumed to be representative samples of two macro-varieties. It is taken as a given that these are not homeogenous varieties, but to the extent that it is appropriate, they will be referred to here as Mexican Spanish and Iberian Spanish. The analyses presented here are naïve to the internal variation inherent to each variety.
regional distributions is statistically significant ($\chi^2 = 122.4, p < 0.001$), however it is difficult to say that either variety is necessarily more conservative or more innovative. The Iberian data clearly exhibit a higher frequency for the innovative \( [\text{haber que} + \text{Infinitive}] \) construction (436 vs. 297 tokens per million), while the rate of \( [\text{tener que} + \text{Infinitive}] \) is higher in the Mexican portion of the corpus, albeit by a comparatively small amount.

### 5.3. The occurrence of \( [\text{tener que} + \text{Infinitive}] \) in the 20\(^{th}\) century

Although neither regional variety appears to favor \( [\text{tener que} + \text{Infinitive}] \) more than the other, the \( \text{tener que} \) construction is clearly the most frequent of the four constructions in the data from each region, occurring nearly twice as often as the next most frequent form, \( [\text{deber (de)} + \text{Infinitive}] \). Regardless of how the data are divided, the \( \text{tener que} \) construction is the predominant form used to encode obligation in these most recent data. The advent of \( [\text{tener que} + \text{Infinitive}] \) was relatively late compared to those of the \( \text{haber de} \) and \( \text{deber (de)} \) constructions, but its frequency rose sharply. The number of tokens of \( [\text{tener que} + \text{Infinitive}] \) expressing obligation was negligible prior to the 17\(^{th}\) century, and the construction quickly gained ground on the alternative variants, making up more than half of the data in the 20\(^{th}\) century (see Figure 5-1 above).

The sharp increase in frequency for \( [\text{tener que} + \text{Infinitive}] \) is not the only evident change the construction has undergone in the transition from the 19\(^{th}\) to the 20\(^{th}\) century. In Chapter 4, the type/token ratio for this construction was discussed, as the ratio of distinct lexical verb types to total tokens had steadily increased over the previous few centuries. However, in the 20\(^{th}\) century, that trend has stopped, as presented in Table 5-6.
Table 5-6. Type/token ratios for [tener que + Infinitive]: 15th century to 20th century
(ratio of number of distinct lexical types appearing in the construction to total number of tokens)

<table>
<thead>
<tr>
<th>Century</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
<th>19th</th>
<th>20th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type/token ratio (# types/total N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15th</td>
<td>0.321 (26/81)</td>
<td>0.324 (105/324)</td>
<td>0.371 (122/329)</td>
<td>0.403 (165/409)</td>
<td>0.472 (226/479)</td>
<td>0.279 (292/1047)</td>
</tr>
</tbody>
</table>

As can be seen in Table 5-6, the type/token ratio has decreased for the first time since the 15th century.41 Despite the fact that the number of distinct lexical verb types appearing within the construction has increased substantially from the 19th to the 20th century (from 226 types to 292), the overall number of tokens of the construction itself has enjoyed an even greater increase, more than doubling from 479 tokens to 1046. This creates a sort of ceiling effect: although the construction continues to be generalized to include even more lexical verb types, the number of possible lexical verbs is finite, while the token frequency of the construction is free to increase beyond even a hypothetical maximum number of possible lexical types. The sheer frequency with which the construction occurs means that individual lexical verbs will be more likely to recur than they did in previous centuries, driving down the type/token ratio (cf. Torres Cacoullos 2002: 310). Rather than aiding in the grammaticalization process as they presumably did in the 15th-18th centuries, recurring lexical verb types are merely evidence of the extremely high frequency of the tener que construction itself.

Thus, a number of lexical verbs appear in the tener que construction quite frequently. Among these are hacer ‘to do, to make’ (96/1047 tokens), ir(se) ‘to go, to leave’ (83/1047 tokens), estar ‘to be’ (38/1047 tokens), and ser ‘to be’ (37/1047 tokens). This is not surprising, as the aforementioned verbs are highly frequent in general. In fact, the single most commonly recurring lexical type, hacer ‘to do’, was one of the lexical verbs that was found to appear repeatedly in the early stages of the tener que construction as it underwent the semantic changes

---

41 See the additional discussion of the effect of very high and very low token frequencies on type/token ratio in Section 3.2.6.
in the 15th-18th centuries that brought it to its current state. However, these modern uses of *hacer* are not necessarily examples of the relative clause construction conveying a sense of possession that was discussed in Chapter 3 (e.g. *tener que hacer* ‘to have (something) to do’):

(142) *lo tengo que hacer, porque es necesario*

‘I have to do it, because it is necessary.’ [20th c.; *MexCult: 26*]

(143) *como estoy trabajando, pues tengo que hacer la comida los sábados*

‘since I am working, I have to make the food on Saturdays’  
[20th c.; *MadCult: 122*]

Examples like those in (142-143) illustrate that the obligation use of the construction has become the primary function of [*tener que + Infinitive*]. Further evidence of the extent to which the *tener que* construction’s meaning has changed is found in the fact that the *tener* ‘to have’ is one of the most frequent lexical types to appear in the construction, occurring 27 times in the 20th century data. That *tener* itself may appear as the lexical infinitive in the [*tener que + Infinitive*] construction is an indicator of the extent to which the lexical (possession) meaning has been bleached from the modal *tener*, as in (144-145).

(144) *Y entiendo que el profesor, pues, una de las cosas que tiene que tener es... que se nazca para profesor.*

‘And I understand that a professor, one of the things that he has to have is...that he is born to be a professor.’ [20th c.; *MadCult: 22*]
5.3.1. Persistence of older meanings

Tokens like (144) exemplify the distinction between the modal meaning of tener in [tener que + Infinitive] and its possession meaning when it appears as the lexical infinitive in the construction. Given examples like these, the sharp increase in frequency of [tener que + Infinitive], and its apparent displacement of the older haber de and deber (de) constructions, it is not difficult to argue that expressing obligation has become the chief employment of [tener que + Infinitive]. This is not to say, however, that its older functions have disappeared altogether. The examples in (146-147) present modern instances of this older use of the tener que construction, the relative clause construction from which the modern modal construction developed (as discussed in Chapters 2 and 3).

(145)  *Estate tranquilo. No va a pasar nada. No tienes que tener miedo de todo.*

‘Be calm. Nothing is going to happen. You do not have to be afraid of anything.’ [20th c.; Davies 2002-]

(146)  *Yo me di cuenta de que estaba delante de un hombre que tenía MUCHAS COSAS que decir.*

‘I realized that I was before a man who had MANY THINGS to say.’ [20th c.; MadCult: 293]

(147)  *un mendigo podía ser más feliz sin saber cada día si tendría ALGO que llevarse a la boca, que un rico en un sumptuoso palacio*

‘a beggar could be happier without knowing if he would have SOMETHING to eat each day, than a rich man in a sumptuous palace’ [20th c.; Davies 2002-]

Although they are now far less frequent than the deontic modal meaning of the construction, relative clause tokens such as these are by no means unusual, as Table 5-7 points out. In fact,
multiple possession-type tokens appear in each of the sources from which the 20th century data are drawn.

Furthermore, the majority of the possession-type tokens in the 20th century are examples of the specific *tener que ver* (con) ‘to have something to do (with)’ which was already relatively frequent in the data prior to the 18th century and persists in the modern language (see section 3.2.4). This particular form of the construction appears to be a sort of offshoot of *[tener que + Infinitive]*, retaining the same meaning it had several centuries ago. In all, 44 of the 73 possession-type occurrences include *ver*, as with (148-149).

(148)   *Con ese viejo no quiero tener NADA que ver.*

‘I do not want to have ANYTHING to do with this old man.’ [20th c.; Pedro: 110]

(149)   *¿Y qué tiene que ver el cáncer con la filosofía?*

‘And what does cancer have to do with philosophy?’ [20th c.; Davies 2002-]

The lexical verb *hacer* ‘to do’ also persists in a few (5 of 73) of the possession-type tokens, exhibiting effectively the same meaning it had in such instances as early as the 16th century (see Section 3.2.3):
(150) como empieces a pensarlo o, o no tienes NADA que hacer es cuando te duele muchísimo más...

‘like if you start to think about it or, or if you don’t have ANYTHING to do is when it hurts a lot more.’ [20th c.; MadCult: 359]

5.3.2. The appearance of innovative modal meaning in [tener que + Infinitive]

Perhaps even more noteworthy than the persistence of older meanings in the data is the apparent advent of an even newer modal sense, one of epistemic modality. That is, [tener que + Infinitive] exhibits a sense of supposition or probability in a number of tokens in the 20th century. In all, 30 tokens conveying this meaning were found in the present corpus.

(151) ponemos el tocadiscos, como los altavoces están al lado del techo, yo creo que la vecina de arriba tiene que bailar

‘we put on the record player, since the speakers are on the side of ceiling, I think the upstairs neighbor must be dancing’ [20th c.; MadCult: 333]

(152) Acaba de estar aquí una señora. ustedes tuvieron que verla salir.

‘A woman was just here. You had to see her leave.’ [20th c.; Pedro: 122]

These innovative occurrences of the tener que construction are less frequent than the persisting possession-type tokens, but they are found in each of the sources in the 20th century corpus. Their appearance is not entirely surprising, given that reference grammars mention this use of the [tener que + Infinitive] construction (Butt and Benjamin 2004: 328, Gómez Torrego 1999: 3353). Furthermore, epistemic modal meaning is a semantic target for erstwhile expressions of obligation and has been attested in numerous languages, including Spanish (Fleischman, 1982; Bybee et al. 1994; Heine and Kuteva 2002).

In fact, the expression of a supposition or probable future is the prescribed use of the [haber de + Infinitive] construction, which has all but vanished from the 20th century data as an
obligation construction. Butt and Benjamin assert that the older obligation meaning of the *haber de* construction is “literary and faintly archaic” in the modern language, while describing a probable future is the primary function of *haber de* (2004: 329). The data presented here suggest that the *haber de* construction had been the dominant modal construction for expressing obligation for centuries prior to the rise of *[tener que] + Infinitive* in that domain (see Chapter 4).

Therefore, it is perhaps to be expected that given its “head start”, the *haber de* construction has become largely limited to expressing epistemic modality as it has continued to grammaticalize. This semantic change in *[haber de] + Infinitive* has evidently been accompanied by a decrease in the construction’s overall frequency. In all, 239 tokens of the *[haber de] + Infinitive* were extracted from the 20th century corpus. Of these 239, only 68 (28%) were interpreted as expressing obligation, and 50 of those 68 were found in the (presumably more conservative) written sources. This is in stark constrast with the 623 tokens of *[haber de] + Infinitive* that were found to express obligation alone in the 19th century.42 The obligation and probability uses of the construction are contrasted in examples (153-154).

Obligation:

(153)  *me has de dispensar, pero yo no maté a Odilón*

‘You’ll have to forgive me, but I did not kill Odilón’ [20th c.; Llano: 29]

Probability:

(154)  *yo no he gritado, Susana. Has de haber estado soñando.*

‘I did not shout, Susana. You must have been dreaming.’ [20th c.; Pedro: 157]

---

42 The 19th century corpus was slightly larger than that of the 20th century (1.1 million words vs. 930,000), but the size difference does not account for the clear drop in the construction’s overall frequency.
A number of the 20th century tokens of [haber de + Infinitive] express a sort of future temporal reference, presumably referring to a future that is probable or inevitable, as in (155-156). The progression from obligation to future reference is another change attested in numerous languages (Bybee et al. 1994; Heine and Kuteva 2002). Indeed, the Spanish synthetic future itself (pensaré en ti ‘I will think of you’ in (18)) includes a grammaticalized form of Latin habēre ‘to have’ that transitioned from expressing obligation to expressing future reference and became suffixed to an infinitive verb form (Fleischmann 1982).

(155) algún día, si Dios quiere, nos hemos de encontrar cara a cara

‘Someday, God willing, we will meet face to face’ [20th c.; Davies 2002-]

(156) aunque nos separemos, siempre pensaré en ti, y siempre ha de pertenecerte mi cariño.

‘even if we separate, I will always think of you, and my heart will always belong to you’ [20th c.; Davies 2002-]

In contrast, the usage of [tener que + Infinitive], to express probability is clearly quite limited. But the 30 tokens expressing probability that were found in the 20th century corpus (as in (151-152) above) are suggestive of its continuing grammaticalization.

5.3.3. Layering of meanings in [tener que + Infinitive]

Importantly for the present investigation, all three of the basic meanings of [tener que + Infinitive] coexist in the 20th century data. While expressing deontic modality is the predominant application of the construction at this stage in its ongoing grammaticalization, the older

---

43 The two examples in (17-19) were taken from a portion of the Davies (2002-) corpus that was not included in the analysis, but are presented here because of their clear illustration of the meaning.
possession meaning and the innovative probability sense both recur in the corpus. This supports the hypothesis that, as Hopper argues, multiple “layers” of meaning will likely exist side-by-side at a given point in the grammaticalization of a form (1991: 22-24). The example in (157) is a paragon of this layering:

(157)  *y con cara de enfermo y de hambriento, debe tener, tiene que tener algo que contar*  

‘and looking sick and hungry, he must have, he has got to have something to tell’ [20th c.; Davies 2002-]

This occurrence of the *tener que* construction illustrates several noteworthy points. First, it demonstrates an epistemic use of the construction, as the speaker is making an assumption based on existing evidence (the subject’s facial expression). In addition, this is a rare occurrence of one token of the construction nested within another. Within the modal construction appears an example of the relative clause-type usage: *tener algo que contar* ‘to have something to tell’. Thus, not only does *tener* ‘to have’ appear as the infinitive in [*tener que* + Infinitive], but that token of *tener* is itself a part of an occurrence of an older form of the *tener que* construction: the [*tener (NP) que* + Infinitive] construction (see especially Section 2.4). This secondary token depicts the lexical meaning of the construction from which the modern modal form developed. Finally, the variation between [*deber (de)* + Infinitive] and [*tener que* + Infinitive] is highlighted. The speaker juxtaposes the two constructions, perhaps intending slightly different meanings, but apparently intending them to be roughly synonymous.
5.4. Multivariate analyses of the variation among the modal constructions: 20th century

The evident interchangability of the *deber* (*de*) and *tener que* constructions in (157) is of particular interest for the present study, given that these two constructions combine to account for 81% of the data in the 20th century (see section 5.2). While their respective uses in the above example might be better characterized as expressing probability, the forms clearly overlap in the domain of obligation:

(158)  *volviendo a la mujer soltera [...] todo mundo le dice: "Sí, te *tienes que* casar; es decir, te *debes* casar. Es tu ideal y todo."*

‘Returning to the single woman [...] everybody says, “yes, you **have to** get married; that is, you **ought to** get married. It is your goal and everything.”’

[20th c.; MexCult: 298]

In the previous chapter, multivariate analyses revealed that several linguistic factors were influential in determining the appearance of [*tener que* + Infinitive] rather than one of the three other modal constructions with which it varies in the expression of obligation. Affirmative polarity, perfective aspect, and 1st person subjects were all found to favor selection of the *tener que* construction. Given the sizable increase in the absolute (token) frequency of [*tener que* + Infinitive] and its newfound predominance in the obligation system in the latest century, it is possible that the magnitude and/or direction of these factors’ effects have changed.

The arrangement of the 20th century corpus analyzed here allows for a comparison of the modern data with the results of the 19th century analyses. It also has the advantage of including data drawn from written and spoken language as well as examples of two distinct varieties of Spanish. To determine what, if any, changes have taken place in the variation among the four deontic modal constructions, the data extracted from the 20th century corpus were coded for the
same linguistic factors as those of the 19th century: polarity, aspect, grammatical person, and direct object presence/placement.

5.4.1. Polarity

In the analyses of the 19th century data, polarity was found to be the most influential factor in accounting for the variation between [tener que + Infinitive] and the three other modal constructions. Affirmative clauses slightly favored the appearance of the tener que construction, while negative contexts strongly disfavored it. This was interpreted as evidence that negative contexts tend to be conservative. The 20th century data are distributed similarly to those of the 19th, as depicted in Table 5-8.

Table 5-8. Distribution of 20th century tokens by polarity (percentage of tokens comprising each variant by polarity in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>deber (de)</th>
<th>haber de</th>
<th>haber que</th>
<th>tener que</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>436 (25%)</td>
<td>49 (3%)</td>
<td>301 (17%)</td>
<td>979 (55%)</td>
<td>1765</td>
</tr>
<tr>
<td>Negative</td>
<td>106 (54%)</td>
<td>11 (6%)</td>
<td>23 (12%)</td>
<td>55 (28%)</td>
<td>195</td>
</tr>
<tr>
<td>Interrogative</td>
<td>6 (19%)</td>
<td>8 (25%)</td>
<td>5 (16%)</td>
<td>13 (41%)</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>548 (28%)</td>
<td>68 (3%)</td>
<td>329 (17%)</td>
<td>1047 (53%)</td>
<td>1992</td>
</tr>
</tbody>
</table>

In the modern data, [tener que + Infinitive] accounts for more than half of the total tokens (53%), but only 28% of the tokens with negative polarity. That is, the tener que construction is considerably less likely to appear in a negative context than it is to appear overall. In contrast, the older deber (de) and haber de constructions are more likely to appear in negative contexts than they are to occur in general (54% and 6% in negative contexts vs 28% and 3% overall, respectively). Given the similar distribution of these data to those from the previous century (see Section 4.6.2), it is expected that negative polarity will again disfavor the occurrence of the tener que construction in the multivariate analyses.
5.4.2. Aspect

Perfective aspect strongly favored the appearance of \([tener \ que + \text{Infinitive}]\) in the 19th century, while imperfective aspect only slightly disfavored it. As depicted in Table 5-9, the \(tener \ que\) construction displays the highest rate of perfective tokens (13%) again in the 20th century, suggesting that this aspect will again favor the selection of \([tener \ que + \text{Infinitive}]\) over the other three forms.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Tense</th>
<th>tener que</th>
<th>haber de</th>
<th>deber (de)</th>
<th>haber que</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfective</td>
<td>Present</td>
<td>590 (48%)</td>
<td>41 (3%)</td>
<td>329 (27%)</td>
<td>267 (22%)</td>
<td>1227</td>
</tr>
<tr>
<td></td>
<td>Imperfect</td>
<td>156 (47%)</td>
<td>14 (4%)</td>
<td>130 (39%)</td>
<td>33 (10%)</td>
<td>333</td>
</tr>
<tr>
<td></td>
<td>(other)</td>
<td>164 (61%)</td>
<td>10 (4%)</td>
<td>72 (27%)</td>
<td>24 (9%)</td>
<td>270</td>
</tr>
<tr>
<td>Perfective</td>
<td>Preterit</td>
<td>137 (85%)</td>
<td>3 (2%)</td>
<td>17 (10%)</td>
<td>5 (3%)</td>
<td>162</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1047 (53%)</td>
<td>68 (3%)</td>
<td>548 (28%)</td>
<td>329 (17%)</td>
<td>1992</td>
</tr>
</tbody>
</table>

On the surface, the distribution of the 20th century does not reveal any particularly noteworthy changes with respect to aspect. However, it is worth noting that \([tener \ que + \text{Infinitive}]\) makes up an overwhelming 81% of the perfective tokens, suggesting a continued favoring effect of perfective aspect. The \(tener \ que\) construction is also the only construction for which the modal verb appears as an infinitive (i.e. \(tener\)). In all, there are 49 infinitive tokens of the \(tener \ que\) construction in the 20th century data, a number of which appear within the periphrastic future construction \([\text{ir} + \ a + \text{Infinitive}]\) ‘to be going to + Infinitive’.

(159) suponiendo que el sol est' aquí y que la tierra est' en esta posición, \(\text{va a tener que atravesar distintos brazos de nuestra galaxia}\)

‘assuming that the Sun is here and that the Earth is here, in this position, it is going to have to cross different arms of our galaxy’ [20th c.; MexCult: 355]
In studies of grammaticalizing forms, a degree of inflectional or morphological impoverishment has been associated with more advanced grammaticalization (Hopper and Traugott 1993: 103-106). Hopper refers to this as decategorialization, describing it as a loss of “discourse autonomy” whereby a form changes from having a “meaning independent of the text […] to having a meaning or function that is relative to the text or some local construction” (1991: 30; and see Section 3.3 here for changes in the status of the erstwhile relative pronon que ‘that which’). In many cases, decategorialization means forms that are more grammaticalized tend to be more limited in the range of tenses and aspects in which they may appear.

For example, the English have to construction has been characterized as a “semi-modal” (i.e. less grammaticalized), rather than a full modal (Brinton 1991; Fischer 1994). This presumably explains the ability to use have to in a periphrastic future construction like that in (20), and the inability to use should in the same construction: compare it is going to have to traverse different arms of our galaxy with it is going to should traverse different arms of our galaxy. Applying this to the appearance of [tener que + Infinitive] in a periphrastic future construction, the lack of such tokens for the other three modal forms could be interpreted as evidence of their being relatively more grammatical than the tener que construction. The multivariate analyses presented below seek to shed further light on the degree of grammaticalization of [tener que + Infinitive].

5.4.3. Grammatical person

The tener que construction was found to be favored by 1st person subjects in the 19th century, while 2nd and 3rd persons disfavored its occurrence. In the 20th century, variants are distributed among the persons much like they were in the 19th. In the previous century, the older haber de and deber (de) constructions were both more limited to 3rd person singular than tener
que, which was more evenly distributed among the persons. Table 5-10 displays the 20\textsuperscript{th} century distribution, suggesting that this general pattern has not changed.

<table>
<thead>
<tr>
<th>Person/number</th>
<th>tener que</th>
<th>haber de</th>
<th>deber (de)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>271 (84%)</td>
<td>6 (2%)</td>
<td>46 (14%)</td>
<td>323</td>
</tr>
<tr>
<td>1PL</td>
<td>132 (73%)</td>
<td>4 (2%)</td>
<td>44 (24%)</td>
<td>180</td>
</tr>
<tr>
<td>2SG</td>
<td>111 (72%)</td>
<td>9 (6%)</td>
<td>34 (22%)</td>
<td>154</td>
</tr>
<tr>
<td>2PL</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3SG</td>
<td>309 (46%)</td>
<td>42 (6%)</td>
<td>322 (48%)</td>
<td>673</td>
</tr>
<tr>
<td>3PL</td>
<td>138 (60%)</td>
<td>7 (3%)</td>
<td>85 (37%)</td>
<td>230</td>
</tr>
<tr>
<td>Total</td>
<td>961 (62%)</td>
<td>68 (4%)</td>
<td>532 (34%)</td>
<td>1561</td>
</tr>
</tbody>
</table>

One difference however, is that the haber de and deber (de) constructions have become even more limited in their distributions, each tending to appear in third-person singular contexts. This would again suggest a more advanced degree of grammaticalization for those two constructions, being that their forms are evidently less evenly distributed among the grammatical persons than that of the more recently developed [tener que + Infinitive]. In contrast, the tener que construction makes up 84\% of the first-person singular tokens century data, which is up from 31\% in the 19\textsuperscript{th}. The portion of first-person plural tokens accounted for by [tener que + Infinitive] has increased from 30\% to 73\% as well. As with the previous century, it is expected that first-person will be selected as a favorable context for the innovative tener que, as its rate of occurrence is higher in first-person contexts than in the data in general; the opposite is true for haber de and deber (de).

\footnote{Tokens for which the subject is usted (2\textsuperscript{nd} person singular, formal) are excluded, as they are semantically 2\textsuperscript{nd} person, but appear in the 3\textsuperscript{rd} person singular form. Non-finite tokens, which are exclusive to [tener que + Infinitive], are also excluded. The haber que construction, which invariably appears with a 3SG subject, is also not excluded.}
5.4.4. Medium and regional variety

To the linguistic factors, two additional extralinguistic factors were included in the 20th century analysis. The first is the medium of the source from which the data were drawn. Tokens were classified as being drawn from written or spoken language, thereby distinguishing the literary data from that of the transcribed interviews. As noted in Section 5.2.1, the spoken data exhibit a considerably higher frequency of \(tener que + \text{Infinitive} \) than the written data, so it is expected that the spoken medium will favor the occurrence of the \(tener que \) construction.

The final factor included in the multivariate analyses is the regional variety. The data for this century were extracted from written and spoken examples of Mexican and Iberian Spanish, as discussed above in Section 5.2.2. Given that the Mexican data exhibit a slightly higher frequency of \(tener que + \text{Infinitive} \) (see Table 5-5), it is hypothesized that the Mexican variety will favor its occurrence.

5.5. Results

As with the 19th century data, the Goldvarb X application (Sankoff, Tagliamonte, and Smith 2005) was utilized to determine the degree to which each of the linguistic and extralinguistic variables influenced the choice of \(tener que + \text{Infinitive} \) over the other variants. Due its sparsity, \(haber de + \text{Infinitive} \) was not included in a paired analysis with \(tener que + \text{Infinitive} \). The 68 tokens of the \(haber de \) construction were, however, included in an analysis comparing \(tener que \) with the other three constructions as a group.

The factor groups were collapsed and/or simplified into binary pairings (perfective aspect vs. imperfective aspect; affirmative vs. negative polarity; 1st person vs. all other persons) as they were in the 19th century analyses to allow for a comparison of the relative degree of effect for
each variable and to maintain consistency with the analysis of the previous century\(^{45}\). The extralinguistic variables, medium and region, were inherently binary.

5.5.1. Individual paired analyses

As with the 19\(^{th}\) century, the specific analysis pairing \([tener que + Infinitive]\) with \([haber que + Infinitive]\) did not include grammatical person as an independent variable, as the \(haber que\) construction appears categorically in 3\(^{rd}\) person singular. Table 5-11 presents the results of the paired analyses with \([deber (de) + Infinitive]\) and \([haber que + Infinitive]\). Factor weights greater than 0.5 favor the occurrence of \([tener que + Infinitive]\); weights less than 0.5 disfavor its occurrence.

---

\(^{45}\) The portion of tokens that was excluded represents only a small percentage of the total data. For example, tokens for which \(tener\) appeared as an infinitive, as with (20) above, made up less than 5\% of the tokens of that construction. Trimming the data this way prevents singletons and knockouts in the distribution of the data that are subjected to the multivariate analyses, and allows the regression analysis to better fit the distribution.
Table 5-11. Results of multivariate analyses pairing [tener que + Infinitive] respectively with
[deber (de) + Infinitive] and [haber que + Infinitive] in the 20th century

<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factor weight</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>vs. deber (de)</td>
<td>vs. haber que</td>
<td></td>
</tr>
<tr>
<td>N = 1342 (Input: 0.696)</td>
<td>N = 1172 (Input: 0.78)</td>
<td></td>
</tr>
<tr>
<td><strong>Aspect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfective</td>
<td>0.86</td>
<td>0.90</td>
</tr>
<tr>
<td>Imperfective</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Range:</strong> 42</td>
<td><strong>Range:</strong> 48</td>
<td></td>
</tr>
<tr>
<td><strong>Polarity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>Negative</td>
<td>0.17</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Range:</strong> 37</td>
<td><strong>Range:</strong> 15</td>
<td></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoken</td>
<td>0.63</td>
<td>0.54</td>
</tr>
<tr>
<td>Written</td>
<td>0.27</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Range:</strong> 36</td>
<td><strong>Range:</strong> 13</td>
<td></td>
</tr>
<tr>
<td><strong>Person</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>0.72</td>
<td>N/A</td>
</tr>
<tr>
<td>All others</td>
<td>0.38</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Range:</strong> 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>0.62</td>
<td>0.54</td>
</tr>
<tr>
<td>Spain</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Range:</strong> 18</td>
<td><strong>Range:</strong> 10</td>
<td></td>
</tr>
</tbody>
</table>

Factor not selected as significant: Object presence

As can be seen in Table 5-11, the linguistic factors are generally more influential in accounting for the variation between [tener que + Infinitive] and the other two constructions. In each pairing, aspect and polarity are the two most influential. The difference between the factor weights for each factor group (the range) is larger for these two variables than for any of the others. Perfective aspect strongly favors the occurrence of the tener que construction, while imperfective aspect slightly disfavors it. Polarity is also influential, but in the opposite direction: affirmative contexts only slightly favor the occurrence of [tener que + Infinitive], while negative contexts strongly disfavor it. The final linguistic variable, grammatical person, displayed the
same pattern it did in the 19th century, with 1st person subjects again favoring \([tener que + \text{Infinitive}]\). All other subjects (combined into a single factor) were somewhat disfavorable.

For the extralinguistic factors, the medium has a larger effect than the regional variety, which was the least influential of the factors included in each analysis. As expected, the spoken data favored the innovative \(tener que\) construction, while the presumably more conservative written medium clearly disfavored it. The regional variety had less of an effect, with Mexican Spanish somewhat favoring the appearance of \([tener que + \text{Infinitive}]\).

5.5.2. Analysis of \([tener que + \text{Infinitive}]\) vs. all other variants

A final comparison of the 20th century data with those of the 19th century is found in the analysis presented below in Table 5-12, grouping together the \(deber\ (de), haber de,\) and \(haber que\) constructions and analyzing them as a single alternative to \([tener que + \text{Infinitive}]\).
Table 5-12. Results of multivariate analysis of factors contributing to the selection of \([tener que + Infinitive]\) over \([deber (de)/haber de/haber que + Infinitive]\): 20th century

Total N = 1721, Corrected mean: 0.53 (51\% tener que)

<table>
<thead>
<tr>
<th>Factor weight</th>
<th>% tener que</th>
<th>Total N</th>
<th>% data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfective</td>
<td>0.86</td>
<td>85%</td>
<td>162</td>
</tr>
<tr>
<td>Imperfective</td>
<td>0.45</td>
<td>48%</td>
<td>1559</td>
</tr>
<tr>
<td>Range: 41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>0.79</td>
<td>80%</td>
<td>474</td>
</tr>
<tr>
<td>All others</td>
<td>0.38</td>
<td>41%</td>
<td>1247</td>
</tr>
<tr>
<td>Range: 41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polarity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative</td>
<td>0.53</td>
<td>54%</td>
<td>1556</td>
</tr>
<tr>
<td>Negative</td>
<td>0.22</td>
<td>22%</td>
<td>165</td>
</tr>
<tr>
<td>Range: 31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoken</td>
<td>0.62</td>
<td>60%</td>
<td>1086</td>
</tr>
<tr>
<td>Written</td>
<td>0.31</td>
<td>37%</td>
<td>635</td>
</tr>
<tr>
<td>Range: 31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factors not selected as significant: Regional variety, object presence

The linguistic variables are again the most influential in accounting for the variation between \([tener que + Infinitive]\) and the alternative variants. When those three constructions are grouped together, grammatical person appears to have a magnitude of effect (Range = 41) equal to that of aspect. However, its influence is likely at least partially inflated by the inclusion of tokens of \([haber que + Infinitive]\) in this analysis. As the \(haber que\) construction appears categorically with 3rd person singular subjects, this could make 1st person subjects appear to be even more favorable to the occurrence of \([tener que + Infinitive]\). Polarity is effectively on equal footing with medium (spoken vs. written), the sole extralinguistic factor group that was selected as significant. In the 19th century, polarity was the single most influential factor group, suggesting
a change in the relative impacts of each of the linguistic variables analyzed here. This diminishing influence of negation is to be expected if negation is associated with conservatism. If \[tener \text{ que } + \text{ Infinitive}\] is no longer as relatively innovative with respect to the \(deber \ (de)\) and \(haber \ de\) constructions (i.e. it has advanced further along its path of grammaticalization), then we would expect polarity to be less influential accounting for the variation between the forms. Aaron (2010) reports similar results in the development of the Spanish periphrastic future construction: negation disfavors the grammaticalizing form, but its effect becomes weaker over time.

However, as with the results of the paired analyses presented in 5.5.1, the linguistic factors have a greater impact in general on the choice between \[tener \text{ que } + \text{ Infinitive}\] and any of the three other obligation constructions. This is perhaps the more interesting result: although the occurrence of the innovative \(tener \text{ que}\) construction is apparently more robust in the spoken language, the general shift from the older \(deber \ (de)\) and \(haber \ de\) constructions is driven more by linguistic variables.\(^{46}\) What is more, those same linguistic variables that were found to have a statistically significant influence on the variation in the obligation system for the 19\(^{th}\) century were again significant in the 20\(^{th}\). And despite the fact that the data were evidently distributed differently in Mexican and Peninsular Spanish (see section 5.2.2), the regional variety does not have a strong enough effect to be selected in the combined analysis.

5.6. Conclusion

A few notable changes in the distribution of the modal constructions discussed here have taken place in the interval between the 19\(^{th}\) and the 20\(^{th}\) centuries. Most importantly, the \([tener \text{ que } Infinitive]\) construction has become the predominant form used to encode obligation,

\(^{46}\) A comparison of the effect of linguistic variables with that of the medium (spoken vs. written language) could be made by analyzing the spoken and written tokens separately (see Section 6.6).
relative to the three alternatives. While it may still be considered relatively innovative with respect to the older \[deber (de) + \text{Infinitive}\] and \[haber de + \text{Infinitive}\], the \textit{tener que}\nconstruction has reached an absolute frequency in the 20\textsuperscript{th} century such that it appears nearly twice as often as the next most frequent construction (\[deber (de) + \text{Infinitive}\]) in the corpus.

The \textit{tener que} construction has become so frequent in fact, that for the first time in the chronological range studied here, its type/token frequency has decreased. That is, until this point, the construction had consistently appeared with more and more lexical verb types with each passing century. While that still holds true for the 20\textsuperscript{th} century, it has evidently reached a plateau of sorts, where the number of distinct lexical types can no longer outpace the sheer rate at which \[tener que + \text{Infinitive}\] occurs. The construction displays the increased productivity that is to be expected with continuing grammaticalization, but its greatly augmented token frequency overshadows that productivity, driving down the type/token ratio for the 20\textsuperscript{th} century (see §5.3).

In contrast, the older \[haber de + \text{Infinitive}\] has receded almost as rapidly as \[tener que + \text{Infinitive}\] has gained ground. In the 17\textsuperscript{th} century, the \textit{haber de} construction was the most frequent of the four variants, but only makes up 3\% of the 20\textsuperscript{th} century data. Rather than encoding obligation, it is now largely limited to epistemic uses instead, commonly referring to suppositions and probable futures.

If the disappearance of \[haber de + \text{Infinitive}\] has left a void, that void has largely been filled by the two newest constructions, \[tener que + \text{Infinitive}\] and \[haber que + \text{Infinitive}\]. But although both of the innovative forms have increased in frequency they have apparently not done so in parallel fashion. The \textit{haber que} construction seems to occupy a sort of niche, given that it appears invariably with a generic 3\textsuperscript{rd} person singular subject. Its emergence coincided chronologically with that of \[tener que + \text{Infinitive}\], but it evidently developed modal meaning via a semantic “detour” through the notion of existence rather than possession (see Chapter 4). While \[haber que + \text{Infinitive}\] is ostensibly based on an erstwhile possession verb, as is \[tener
The former construction does not appear to have arrived at its modern modal via the same path as the latter. The tener que construction clearly passed gradually through an intermediate stage of possession with implied responsibility on its way to acquiring deontic modal meaning. Further investigation of the [haber que + Infinitive] construction would be required to confirm this (see Section 6.6 for more discussion of this).

This brings up an important point: not all grammaticalizing forms necessarily proceed through the process at the same rate or follow the same semantic trajectory. Even with two forms that have followed similar paths (from effectively the same semantic source to roughly synonymous targets), the haber que and tener que constructions have not plotted identical courses as they have done so. Despite sharing a start and endpoint with [tener que + Infinitive], the haber que construction passed through a distinct intermediate stage. It could therefore be argued that its grammaticalization is not, entirely analogous to that of [tener que + Infinitive], as is suggested by Yllera (1980: 109-110). The haber que construction, actually exhibits signs of being relatively ‘more grammaticalized’ than the tener que construction: it is more limited in its distribution among grammatical persons (confined to 3rd person singular) and the modal verb itself is almost invariable in its formal appearance (it typically occurs as hay).

A final observation to be made about the 20th century data concerns the linguistic variables that were found to be influential in accounting for the variation among the four modal constructions studied here. The same factors (grammatical person, polarity, and aspect) were operative in the 19th century, and the direction of their effects is what was anticipated. Indeed, continuity of the factors that condition synchronic variation as part of an ongoing change has been reported in several studies of grammaticalizing forms (Poplack and Malvar 2007; Copple 2009; Aaron 2010; Torres Cacoullos 2011). This finding attests to the appropriateness of applying variationist methodology to the analysis of a gradual change such as this.
But that is not to say that the status of the four modal variants as a collective group in the 20th century is identical to that of the previous century. As has been discussed, [haber de + Infinitive] has effectively vanished from the domain of obligation, and [haber que + Infinitive] could be considered to be something unique because of its formal idiosyncrasies. What was a system of four more or less equivalent constructions has been reduced to a “two-horse race” with an apparent winner: [tener que + Infinitive] and [deber (de) + Infinitive] combine to account for 81% of the 20th century data, and the former is now nearly twice as frequent as the latter.

In these latest data, the tener que construction shows itself to have thoroughly acquired modal meaning. Although its older lexical (possession) meaning persists in the 20th century, that particular usage is not nearly as frequent as the modal usage. And while a new epistemic application of the construction has emerged, the deontic meaning is clearly dominant. It is tempting to speculate that [tener que + Infinitive] will continue to grammaticalize and potentially emulate [haber de + Infinitive], eventually being applied most commonly to expressing suppositions. But it is without doubt that the tener que construction has arrived at the endpoint of the specific path from possession to obligation.
Chapter 6

Conclusions: Gradual development of [tener que + Infinitive] and evidence of grammaticalization processes

The research presented here endeavored to trace the diachronic development of the Spanish [tener que + Infinitive] construction. By compiling a corpus of written texts and transcribed interviews, the target construction was followed from its earliest occurrences in the 12th century to its manifestation in the modern language. Using measurable evidence, the semantic and formal changes that brought about the 20th century modal version of [tener que + Infinitive] were assessed and described.

In addition to the case study of this particular construction, the results of the investigation were considered in the larger context of crosslinguistic tendencies in grammaticalization. The existing literature on grammaticalization presents generalizations about the processes that are argued to be at work as lexical items take on grammatical meaning. Taking grammaticalization theory as a hypothetical perspective, the present study finds evidence of a number of those proposed inherent processes.

6.1. Motivation for the study

As noted in Chapter 1, it was noted that the extant literature makes observations about phenomena that coincide as a form grammaticalizes. Crosslinguistic tendencies of semantic change, in particular the attested transition from a lexical meaning that encodes possession to a modal meaning that denotes obligation, form the foundation for this research. A number of authors have corroborated this change (Denning 1987; Bybee, Perkins and Pagliuca 1994; Heine and Kuteva 2002). But a thorough analysis of this type of change, based on quantitative evidence, is missing from the discussion of the topic heretofore. In addition to contributing a systematic
approach to a specific semantic development, the results of this study provide material testimony that confirms a number of predictions made in the grammaticalization literature, sheds doubts on some particular claims about Spanish obligation expressions, and offers a fresh view of the configuration of factors shaping the domain of obligation. A number of individual factors operative in the grammaticalization of [tener que + Infinitive] and in the variation between it and three other modal constructions have been identified and quantified.

6.2. Hypotheses regarding semantic changes

Grammaticalization as a general phenomenon has been argued to incorporate multiple component processes which individually affect various aspects of a form and its meaning (Hopper and Traugott 1993). Of particular importance to the evolution of the [tener que + Infinitive] construction is its semantic development, an aspect of grammaticalization that has been thoroughly discussed. And perhaps chief among the distinct elements and types of semantic change active in grammaticalizing forms are the notions of “bleaching” and “abstraction” (Heine, Claudi and Hünnemeyer 1991: 39-41). Semantic bleaching assumes that a grammatical(izing) form’s meaning is already included as part of the inherent semantic structure of its lexical source (Givón 1973). That is, as a lexical form grammaticalizes, specific components of its meaning are stripped away, leaving only that crucial component necessary for its expression of grammatical meaning. In the case of [tener que + Infinitive], this would imply that some sense of modal meaning is present in the earliest occurrences of the construction. Consider the example in (160):
(160) \textit{nin tenemos pan que comer otra cosa ninguna} \\

neither have.PRS.1PL bread REL eat.INF nor other thing neither \\

‘We have neither bread (which) to eat nor any other thing.’

[13\textsuperscript{th} c.; \textit{GEI}: II, 179]

In (160), one of the earliest manifestations of the 	extit{tener que} construction is depicted. \textit{Tener} is clearly still a lexical verb in this instance, denoting the possession of a concrete object. Still, it is argued here that the relative clause construction carries with it a sense of purpose and/or intention to perform the action of the infinitive \textit{comer} ‘to eat’ (see Chapter 2). Based on that assumption, there is already in the first occurrences of [\textit{tener que} + Infinitive] a hint of reference to a proposition not yet completed, but projected into a necessary or probable future (i.e. denoting obligation and/or probability). This portion of the meaning, albeit tangential, would allow for conceiving of the transition from the construction’s lexical meaning to its modal meaning as a reduction of its semantic content to only this inherent aspect.

However, the inclusion of the relative clause \textit{que comer} ‘(which) to eat’ is crucial to this argument. The verb \textit{tener} ‘to have’ itself does not include a modal component to its meaning; that element is only present within the complete [\textit{tener que} + Infinitive] construction. Given this complication, the bleaching argument does not account for the semantic change undergone by \textit{tener} in [\textit{tener que} + Infinitive]. Indeed, Heine et al. assert that the notion of semantic bleaching is “inadequate as a descriptive or explanatory parameter of grammaticalization” (1991:41).

An alternative description of the change is one of “abstraction” away from a concrete lexical source. That is to say that the target meaning for a grammaticalizing form is generally more abstract than its source (Heine et al 1991: 41-43). Hopper and Traugott describe the early stages of the transition this way by characterizing them as “redistribution or shift, not a loss, of meaning” (1993: 88). The change exemplified by [\textit{tener que} + Infinitive] provides support for this
conceptualization of semantic change in grammaticalization. The example in (161) includes the lexical infinitive *dar* ‘to give’, which contributes a sense of responsibility to repay what is presumably a debt.

(161)  
*Yo no tengo dineros que os dar por la llave*

‘I do not have money to give you for the key’ [16th c.; Lazarillo: 26]

This illustrates the “shift” mentioned by Hopper and Traugott. It is the specific lexical verb *dar* and the contextual allusion to a required payment inherent to the object *dineros* ‘money’ that allow for the meaning of *tener* ‘to have’ to transition to a semantically adjacent notion. That is, rather than mere possession, the occurrence of the construction in (161) overlays possession with a sense of responsibility. The addition of responsibility to the overall semantic content of the construction creates the potential for bridging contexts, as discussed in Chapter 3 (Heine 2002: 84-85). In a bridging context, there is ambiguity between the two general meanings (possession and responsibility/intention):

(162)  
*_…cuan camino quisiere ahorrar caminos y trabajos para llegar a la inaccesible cumbre del templo de la Fama, no tiene que hacer otra cosa sino dejar a una parte la senda de la poesía, algo estrecha, y tomar la estrechísima de la andante caballería*_

*…if you wish to spare yourself odysseys and labors in reaching the inaccessible summit of the temple of fame, you have nothing to do but to leave aside the somewhat narrow path of poetry and take the even narrower one of knight-errantry [17th c.; Quijote: 780]*

Tokens like (162), found in the 17th century corpus, set the stage for conventionalization of the newly added meaning, isolating it from the notion of possession. Conventionalization of the
responsibility meaning results in a usage for which the original lexical content of tener ‘to have’ is absent, leaving only obligation:

(163)  *Eran hombres ocupados que tenían que madrugar.*

‘They were busy men who had to rise early.’ [19th c.; Regenta: 262]

Thus, the semantic changes present in the evolution of [tener que + Infinitive] are best predicted by Hopper and Traugott’s assertion that semantic bleaching “pertains almost exclusively to the late stages of grammaticalization” (1993: 93). Semantic content is ultimately lost, but only after lexical tener appears in a specific construction, and that construction permits the addition of new semantic content that was not originally present.

Additional predictions about the nature of semantic change in grammaticalization are presented by Hopper (1988: 22), who suggests that as new layers of meaning emerge in a grammaticalizing form, the older layers may remain. The tener que construction supports this claim, again demonstrating that semantic content is not necessarily lost until the later stages of the development, and even then, older elements of meaning persist. This is particularly true in the 20th century data, in which the obligation meaning of [tener que + Infinitive] coexists with its earlear possession sense (164) and an even newer expression of probability (165).

(164)  *me di cuenta de que estaba delante de un hombre que tenía muchas cosas que decir.*

‘I realized I was before a man who had a lot of things to say’ [20th c., MadCult: 293]

(165)  *Acaba de estar aquí una señora. Ustedes tuvieron que verla salir.*

‘A woman was just here. You had to see her leave.’ [20th c., Pedro: 122]
6.3. Changes in frequency, distribution, and productivity: quantitative and chronological evidence

In addition to arguments regarding the nature of semantic change in grammaticalization, much attention has been paid to the rate at which this change occurs. It has been argued that grammaticalization takes place gradually and that as a form grammaticalizes, its evolution is accompanied by an increase in that form’s frequency as it generalizes throughout the grammar (Hopper and Traugott 1993; Bybee 2010, inter alia). The present study allows for a particularly detailed test of these hypotheses.

By conducting a century-by-century analysis of [tener que + Infinitive], the results of the research discussed in Chapters 2-5 capture the beginning, middle, and latest stages of the construction’s grammaticalization. The benefit of this approach is that those three general stages are presented not as discrete, serial periods in the construction’s evolution, but instead as overlapping, heterogenous epochs. In the following sections, the quantitative results are summarized and compared to the predictions made by the grammaticalization literature.

6.3.1. 12th-14th centuries

In this earliest era, the [tener que + Infinitive] construction is extremely infrequent. Representing less than 1% of all of the occurrences of the verb tener in the corpus for the 12th-14th centuries, the target construction appears only 14 times in a corpus of 1,150,700 words (~12 tokens per million words). All of these occurrences indicate possession, as in example (160) above. The extremely low rate of appearance of the construction in this time period is crucial to the analysis: this means that the origin of the modern modal construction is effectively captured by the chronological range selected for analysis.
The verb *tener* ‘to have’ continues to vie for the semantic domain of possession with the earlier *haber* ‘to have’ in these first three centuries. This suggests that it is safe to assume the *tener que* construction did not appear with any appreciable frequency any earlier than the 12th century. What is more, a number of these early tokens exhibit a concrete object intervening between *tener* and the relative pronoun *que*, another indication that the construction had only just begun to coalesce as a conventionalized unit:

(166) *dond tengo yo carnes que dar a tan mucha yent como esta*

‘Where do I have meat *which* to give to so many people as these?’

[13th C.; GE1: II, 696]

### 6.3.2. 15th-18th centuries

The second time period analyzed here represents a vital turning point not just for the semantic evolution of the *tener que* construction, but also the frequency with which it occurs. By the 18th century, the construction appears more than 100 times per million words in both the original corpus compiled here and the supplemental data drawn from the *Corpus del Español* (Davies 2002-). This is a marked increase, making [*tener que* + Infinitive] some ten times more frequent than it was prior to the 15th century.

Contributing to the rising frequency of the construction is its increasing productivity, i.e. its use with a growing number of distinct lexical verb types. 26 different lexical verbs appeared as the infinitive component of [*tener que* + Infinitive] in the 15th century corpus, compared to 165 in the 18th century. As the construction becomes increasingly generalized for use with a wide range of verb types, it is still worth noting that *ver* ‘to see’ and *hacer* ‘to do’ have been incorporated into conventionalized forms of the construction (see sections 3.2.3 and 3.2.4, respectively). These
two infinitives are generally frequent, and presumably aid in driving up the rate of occurrence for 
[tener que + Infinitive]. Ver and hacer alone account for a combined 21% of the total tokens of 
the construction in this time period.

Despite the large leaps in productivity and frequency in each subsequent century from the 
15\textsuperscript{th} to the 17\textsuperscript{th}, the tener que construction remains relatively infrequent when compared to its 
antecedents in the obligation system, [haber de + Infinitive] and [deber (de) + Infinitive]. It is not 
until the end of this intermediate epoch that [tener que + Infinitive] makes up an appreciable 
portion of the combined tokens of the four modal constructions considered here to be members of 
the system. In the 18\textsuperscript{th} century, [tener que + Infinitive] accounts for 8% of the data, compared to 
less than 1% for the 17\textsuperscript{th} century.

6.3.3. 19\textsuperscript{th} and 20\textsuperscript{th} centuries

The final two centuries included in the chronological range exhibit another leap in the 
overall frequency of [tener que + Infinitive] and major changes to the distribution of the four 
obligation constructions. In the 19\textsuperscript{th} century, the tener que construction continues its upward 
trajectory, reaching a frequency of 427 tokens per million words. This more than triples its rate 
for the 18\textsuperscript{th} century. In addition, 226 distinct lexical verbs appear in the 19\textsuperscript{th} century data, a 37% 
increase from the previous century.

What is perhaps most striking is that tokens of [tener que + Infinitive] account for 22% of 
the combined occurrences of the four obligation constructions. As the tener que construction 
enjoys this surge in its relative frequency, it evidently encroaches on the territory occupied by the 
preexisting constructions. This is especially true for [haber de + Infinitive], which recedes from 
making up 61% of the combined data in the 18\textsuperscript{th} century to only 29% in the 19\textsuperscript{th}. 
In the 20th century, these trends are largely continued, but the expansion of [tener que + Infinitive] is amplified even more. The overall frequency of the construction nearly triples again, as it now occurs 1122 times per million words (more than once per thousand words in the corpus). A threshold has now been crossed, as the type/token ratio for the tener que construction actually decreases for the first time in the analysis. That is, although the number of distinct lexical verb types appearing in the construction has increased again (up from 226 to 292), it is outpaced by the sheer rate of tener que tokens. In the 19th century, there was one distinct lexical verb type for every two tokens. But in the 20th, that ratio has fallen to 1:3.

Once again, [tener que + Infinitive] makes up an even larger portion of the combined data, as it now accounts for 53% of all tokens of the four constructions (and an even higher 61% in the spoken data). As with the 19th century, this change is accompanied by a further decline in the occurrence of [haber de + Infinitive], which has plummeted to making up only 3% of the data (down from 61% just two centuries before). By the end of the eight-century period described here, the tener que construction has achieved a dominant status with respect to the alternative obligation constructions; it occurs nearly twice as often as [deber (de) + Infinitive], the next most frequent form. As with the hypotheses regarding qualitative semantic changes, the quantitative evidence confirms the predictions made in the grammaticalization literature. The [tener que + Infinitive] construction has been catapulted to an extremely high frequency, and its ascent has been accompanied by a generalization of the construction for use with a broad range of lexical verb types.

Finally, by dividing the complete chronological range into individual centuries, the present research establishes that the advancement of tener que is a gradual progression. In its early stages, [tener que + Infinitive] slowly increased in frequency as it evolved from a specific and rare construction expressing possession to a somewhat conventionalized usage which included a sense of responsibility. This change alone took several hundred years, and it was not
until roughly the 16th century that the construction had established a consistent and substantial rate of occurrence. This same gradualness is evident in its later development: rather than a sudden, abrupt surge in the construction’s frequency following the emergence of the earliest tokens conveying obligation, [tener que + Infinitive] enjoys steady gains from the 17th century onward. Still the rate of change was not constant, as the construction enjoys substantial leaps in its frequency relative to the alternative forms in the last few centuries, especially between the 19th and 20th, as depicted in Figure 6-1.

Figure 6-1. Relative frequency of [tener que + Infinitive] with respect to three alternative obligation constructions ([haber de / deber (de) / haber que + Infinitive])

6.3.4. Multivariate analyses

To provide detailed insight into the linguistic variables that influence the variation between [tener que + Infinitive] and the three alternative obligation constructions, as set of multivariate analyses were conducted. Again, a number of predictions made in the extant
literature were supported. Evidence for the suggestion by Traugott (1989) that the occurrence of a relatively less grammaticalized form should correlate with less subjective contexts was found in the generally favorable impact of 1st person subjects on the occurrence of [tener que + Infinitive]. Givón (1978) provides an additional hypothesis regarding polarity by asserting that negatives are generally more marked and more limited in distribution. This was also evident in the data, as negative contexts disfavored the tener que construction, instead preferring the older (more conservative) haber de and deber (de). A final prediction was drawn from Klein-Andreu’s finding that an older, presumably more conservative form may become increasingly relegated to background events. Assuming that perfective contexts are relatively foregrounded with respect to imperfective aspect (Bybee, Perkins and Pagliuca 1994), this third hypothesis was confirmed in the results: perfective aspect favored the innovative [tener que + Infinitive].

6.4. Implications for theoretical descriptions of grammaticalization processes

Given that a number of the findings reported here were effectively predicted by authors in the existing literature on grammaticalization, some of the results are not altogether surprising. However, the contribution of the present research lies in its methodological approach. Investigating grammaticalization from the perspective of a specific case study rather than as a general phenomenon, the quantitative and chronological detail outlined in the Chapters 2-5 provide concrete evidence of the operative processes inherent to grammaticalization.

By systematically tracing the trajectory of the [tener que + Infinitive] construction, the analyses conducted here provide measurable correlates of the mechanisms subsumed under grammaticalization. For example, rather than simply stating that a grammaticalizing form increases in frequency by comparing its start- and endpoints, nine centuries worth of data demonstrate exactly how that increasing frequency is manifest. Examining the evolution this way
allows us to see the tangible correlates of gradually increasing frequency: the consistent upward trajectory of the absolute frequency of its occurrence, the application of the construction to ever larger numbers of lexical verb types, and the eventual decrease in type/token ratio as the absolute frequency outpaces the number of distinct lexical verbs.

Another advantage of the methods adopted here is that the absolute frequency of \([tener que + \text{Infinitive}]\) can be compared to its frequency relative to alternative constructions. Hopper implicitly acknowledges the importance of alternating forms to grammaticalization in his discussion of layering (1991: 22-24). Viewing a form undergoing change as just one component of a complex, variable system is a notion proposed decades ago by researchers advocating a theory of historical change based on empirical observations (see especially Weinreich, Labov and Herzog 1968). But the benefits of this perspective have only recently begun to be explored in the grammaticalization literature (Poplack 2011).

Juxtaposing the specific \((tener que)\) with the general (the three other obligation constructions) places the innovative form in a richer context and compares its own increasing rate of occurrence with its incursion into the functional domain of expressing obligation. Furthermore, the subsequent variation and semantic overlap described between the four forms can be taken as tangible evidence of what Sankoff refers to as neutralization in discourse (1988b). This provides insight into how the grammaticalization of one form can affect other forms that have grammaticalized along a similar path.

In addition to investigating claims about the grammaticalizing construction’s frequency, the present research found quantitative evidence of the nature of semantic layering and persistence of older meanings within the \([tener que + \text{Infinitive}]\). As mentioned above, the older relative clause (possession) instance of the construction persists in the 20th century, coexisting with both the obligation meaning and the even more innovative expression of probability (see section 5.3). It is important to note, however, that evidence of layering does not mean that the
three basic meanings occur with equal frequency. The obligation sense of the construction is without a doubt the dominant meaning, appearing 1047 times in the 20th century data. The conservative possession usage only appears 73 times (6% of the total 1150 tokens), and the probability usage is even less frequent, occurring only 30 times (3%). Again, the detailed evidence afforded by a systematic, quantitative analysis clarifies the nature of the semantic layering. The meanings do coexist in the modern language, but as [tener que + Infinitive] has grammaticalized, what was the innovative meaning (obligation) has become the primary meaning of the construction.

6.5. The path from possession to obligation

One of the most revealing elements of the research presented here is the distribution of the four constructions included in the Modern Spanish obligation system. Along with [tener que + Infinitive], the three alternative forms with haber de, deber (de), and haber que combine to form a group of semantically related constructions, each of which is derived, at least indirectly, from a possession verb. For that reason, the history of Spanish provides a particularly fortuitous opportunity to compare the grammaticalization paths taken by each of these constructions.

Interestingly, despite the similarities between their lexical sources and their roughly synonymous target meanings, the present study suggests that the paths they took between these two points were not identical. What is more, even the [tener que + Infinitive] construction itself did not trace a linear succession from possession to obligation, another finding with important implications for a theoretical account of how grammaticalization of a construction proceeds. The two prefabs tener que hacer ‘to have (something) to do’ and tener que ver ‘to have to do with’

47 The deber (de) construction is less transparently linked to the notion of possession, but the verb deber ‘should, ought to’ can be traced etymologically to a specific usage of the Latin habère (Vaan 2008: 162), which is the source of Spanish haber, the erstwhile possession verb later replaced by tener ‘to have’.
emerged as offshoots from the general construction, and both played roles in its evolution (see Sections 3.2.3 and 3.2.4). Much like a terminal branch on a family tree, the former disappeared from conventional use, while the latter remains in use in the modern language. Both of these particular forms of the tener que construction appear to have played important roles in the development of the construction. Thus, the data presented here provide evidence of lexical diffusion as the use of tener que spreads from specific, highly frequent lexical infinitive types to a broader range of types including less frequent infinitives⁴⁸. The change did not happen abruptly and moreover lends support to the argument that rather than considering individual manifestations of tener que (e.g. tener que hacer and tener que ver) as separate constructions, it is best to view them all as a single construction (or at the very least a closely related family of constructions). Without acknowledging the interrelatedness of specific forms with the general tener que construction, the explanation of how tener que evolved is lost.

What is more, the oldest form of the construction, for which the interpretation is one of possession, persists in the modern language. It is far less frequent in the data than the modal form of the tener que construction, but such tokens do appear. Indeed, the example in (167) from the 20th century includes the same lexical infinitive (comer ‘to eat’) as was found in some of the earliest tokens in the corpus from the 13th century.

(167) …diría que no tenía yo qué comer allí

‘…I would say that I did not have anything to eat there’ [20th c.; MexPop: 209]

Although the semantic changes that took place in the evolution of [tener que + Infinitive] were found to be largely analogous to those suggested for the history of English have to, existing studies of the grammaticalization of have to include comparatively little quantitative detail

Furthermore, the possibility for differences between seemingly equivalent grammaticalizing forms is readily apparent when comparing the development of \([tener \ que + \text{Infinitive}]\) with that of \([haber \ que + \text{Infinitive}]\). Their forms are ostensibly almost identical: an erstwhile possession verb combined in a particular context with a reflexive pronoun and a lexical infinitive. What is more, the evidently began to emerge as expressions of obligation around the same time. However, as Figure 6-2 illustrates, the construction with tener quickly outpaced the other.

![Figure 6-2](image)

**Figure 6-2.** Token counts for \([tener \ que + \text{Infinitive}]\) and \([haber \ que + \text{Infinitive}]\) from the 18\textsuperscript{th} to the 20\textsuperscript{th} century

How then, do two constructions with seemingly identical sources, congruent structures, following the same path of grammaticalization in the same language at the same time, not perfectly mirror each other along the way? As was suggested in Chapters 4 and 5, it appears that the \([haber \ que + \text{Infinitive}]\) construction took a different route than did \([tener \ que + \text{Infinitive}]\),
eventually acquiring its modal meaning via the expression of existence with *hay* ‘there is/there are’. Given this evidence and the fact that the frequency of [*haber que* + Infinitive] does not increase as much as that of the *tener que* construction, it appears that while it is possible to generalize about recurring paths of grammaticalization as crosslinguistic tendencies, that does not mean the process unfolds the same way in each instance.

### 6.6. Future investigation

Along with the specific contributions of the results presented here, the research outlined in Chapters 2-5 provides ample opportunity for continued examination of the development of deontic modal meaning from the notion of possession. One area that bears investigating more deeply is the potential role played by extralinguistic factors in conditioning the variation between the four modal constructions as well as the grammaticalization of [*tener que* + Infinitive]. Although the linguistic medium (spoken vs. written) was not found to be as influential as the linguistic factors studied here, the higher frequency of occurrence for *tener que* in the interviews (Section 5.2.1) is suggestive of a correlation between prestige and approximating prescriptive norms and the choice of one of the other constructions over [*tener que* + Infinitive].

The influence of medium on the grammaticalization and variation can be analyzed more thoroughly through a specific comparison of written and spoken data. Furthermore, the issue of prestige can be approached via data from theatrical works, not unlike those discussed in Chapter 4. A detailed study of individual characters in the plays may reveal an attempt by the author to portray authentic speech through socially-stratified speakers (e.g. “working-class” characters vs. aristocratic characters); a potential by-product of this would be differing rates of use for the alternating modal constructions.
Perhaps even more replete with opportunities for continued examination of the topic are those obligation constructions themselves. In the history of the Spanish language, the lexical verb *haber* ‘to have’, in two separate constructions and at two separate times, was applied to the expression of obligation. *Haber* was replaced as a possession verb by *tener* ‘to have’, which itself acquired modal meaning in a specific construction. The analyses outlined here were aimed at the latter form, but the former provides an additional avenue of inquiry. A thorough study of the *haber que* construction and its predecessor with *haber de* would shed light on differences between how these two forms have grammaticalized and what was found for [*tener que* + Infinitive].

The construction with *haber de* receded from the present data just as quickly as *tener que* rose to prominence. And despite being almost exclusively utilized to encode epistemic modality, a domain on which the innovative *tener que* construction is evidently only beginning to encroach, *haber de* has nearly vanished in the 20th century. The decline of *haber de* constrasts with *deber* (*de*), which continues to enjoy a relatively high frequency (with both deontic and epistemic modality) in the modern language.

A systematic, diachronic approach like that used here is well-suited for similar investigations of the three obligation constructions that were considered alternatives to [*tener que* + Infinitive] in this study. This is perhaps most applicable to the [*haber que* + Infinitive] construction, whose emergence was evidently contemporaneous with that of [*tener que* + Infinitive], but appears to have developed in a slightly different way, not entirely identical to the evolution of *tener que*.

Such research can provide additional evidence of the manner in which a possession verb develops deontic modal meaning. Moreover, the results of these additional studies can be compared to what has been presented here, confirming or refuting the hypothesis that although
two forms may grammaticalize along the same basic cline, they will not necessarily do so in the
same manner or even at the same rate.
References

Aaron, Jessi Elana. (2010). Pushing the envelope: Looking beyond the variable context.


Brinton, Laurel J. (1991). “The origin and development of quasimodal ‘have to’ in English”.


Thompson, S. A. (1998). A discourse explanation for the cross-linguistic differences in the
grammar of interrogation and negation. In Anna Siewierska and Jae Jung Song (eds.),

*Case, typology and grammar: In honor of Barry J. Blake*, 309–341. Amsterdam: John
Benjamins.

Torres Cacoullos, Rena. (2002). *Le: From pronoun to verbal intensifier. Linguistics* 40.2:
285-318.

--------------. (2006). Relative frequency in the grammaticization of collocations:
nominal to concessive *a pesar de*. *Selected proceedings of the 8th Hispanic Linguistics
Symposium*, Timothy L. Face and Carol A. Klee (eds.), 37-49. Somerville, MA:
Cascadilla Proceedings Project.

--------------. (2009). Variation and grammaticisation: The emergence of an aspectual
opposition. *Studies in Language Variation: European Perspectives II. (ICLaVE 4,
University of Cyprus, 17-19 June 2007)*, 215-224.Stavroula Tsiplakou, Marilena


--------------. (2012). Grammaticalization through inherent variability: The development

middle marking in Spanish *subir(se)* ‘go up’ and *bajar(se)* ‘go down’. *Journal of
Pragmatics* 40:1455-1477.

Traugott, Elizabeth Closs. (1982). From propositional to textual and expressive meanings: Some
semantic-pragmatic aspects of grammaticalization. *Perspectives on historical linguistics*,

--------------. (1989). On the rise of epistemic meanings in English: An example of
-----------------


Appendix

Corpus of texts and transcribed interviews

12th-14th centuries (Chapter 2)

[Cid] (circa 1140) = Anonymous. *Cantar de mio Cid. Texto, gramática y vocabulario*,
[30,000 words]


[EE1] (1270-1284) = Alfonso X, *Estoria de España*. The Electronic Texts and Concordances of
the Prose Works of Alfonso X, El Sabio, Lloyd Kasten, John Nitti, and Wilhemina
Jonxis-Henkemens (eds.). Hispanic Seminary of Medieval Studies. [279,530 words]

Castalia, 1982. [34,000 words]

Planeta. [58,000 words]

[Lucanor] (1335) = Don Juan Manuel, *El conde Lucanor o Libro de los enxiemplos del
[79,200 words]
15\textsuperscript{th}-18\textsuperscript{th} centuries (Chapter 3)


19\textsuperscript{th} century (Chapter 4)

Novels:

[Regenta] (1884) = Leopoldo Alas. \textit{La Regenta} (Tomo I), Gonzalo Sobejano ed., Madrid:
Editorial Castalia, 1981. [146,000 words]

[Doña Perfecta] (1876) = Benito Pérez Galdós. Doña Perfecta, New York:
American Book Company, 1903. [65,500 words]

[Los pazos de Ulloa] (1886) = Emilia Pardo Bazán. Obras completas de Emilia Pardo Bazán
Tomo III: Los pazos de Ulloa, Madrid: Alrededor del Mundo, 1886. [83,800 words]

[Pepita Jiménez] (1874) = Juan Valera. Pepita Jiménez, New York: D. Appleton y Compañía,
1911. [56,500 words]

Theatrical works:

[Acertar errando] (1832) = Ventura de la Vega. Acertar errando, o El cambio de diligencia,
Madrid: Imprenta de Repullés, 1832. [14,000 words]

[La gran comedia] (1884) = Enrique Gaspar. La gran comedia, 1884. [16,000 words]

[Los presupuestos] (1852) = Pablo Alonso de la Avevilla. Los presupuestos, Madrid:
C. González, 1852. [16,000 words]

[Amor de padre] (1849) = Francisco Martinez de la Rosa. Amor de padre, 1849. [18,700 words]

[¡El siete!] (1889) = Julio Cuevas. ¡El siete!, Madrid: R. Velasco, 1889. [5000 words]

[No hay mal que por bien no venga] (1868) = Manuel Tamayo y Baus. No hay mal que por bien
no venga. In Obras completas, Madrid: Fax, 1947. [15,500 words]

[Las circunstancias] (1867) = Enrique Gaspar. Las circunstancias. Madrid: José Rodríguez,
1867. [15,000 words]

Additional document:

[DLNE] (1800-1816) = Concepción Company Company. Documentos lingüísticos de la Nueva
España. Altiplano central, Mexico: Universidad Nacional Autónoma de México,
1994. [21,000 words]
Supplemental texts (from Davies (2002-)):

[El antecristo] Francisco Navarro Villoslada. [75,180 words]

[La Princesa de Viana] Francisco Navarro Villoslada. [36,748 words]

[Compendio de la vida de San Alfonso María de Ligorio] Francisco Navarro Villoslada. [17,509 words]

[La mujer de Navarra] Francisco Navarro Villoslada. [6309 words]

[Doña Blanca de Navarra, crónica del Siglo XV: intitulada Quince días de reinado]
Francisco Navarro Villoslada. [159,114 words]

[El caballero sin nombre] Francisco Navarro Villoslada. [23,500 words]

[El copo de nieve] Ángela Grassi. [78,261 words]

[La mujer de todo el mundo] Alejandro Sawa. [37,843 words]

[Genio e ingenio del pueblo andaluz] Fernán Caballero. [99,104 words]

[Manual del viajero y guía de los forasteros en Valencia] Vicente Boix. [64,884 words]

[Historia de muchos Pepes] Francisco Navarro Villoslada. [49,769 words]

20th century (Chapter 5)

Transcribed interviews:


[MexPop] = Lope Blanch, Juan M. (1976). El habla popular de la ciudad de México:
Materiales para su estudio. Mexico City: UNAM. [177,000 words]

Mexican authors:


Spanish authors:


VITA

Joseph Bauman

Education:
The Pennsylvania State University, University Park, Pennsylvania USA
2013 Ph.D. (Spanish Linguistics) - Defense date: April 12th, 2013
   Dissertation director: Rena Torres Cacoullos
   Dissertation: “From possession to obligation: modal grammaticalization and variation”
2009 M.A., Spanish Linguistics
2004 B.A., Spanish; concentration in Linguistics
2004 B.A., Integrative Arts; concentration in Metal arts

Teaching experience:
2006 – Present; Penn State University – University Park, Pennsylvania USA
   SPAN 002 – Elementary Spanish II
   SPAN 003 – Intermediate Spanish
   LING 100 – Foundations of Linguistics
   SPAN 215 – (Intensive) Introduction to Spanish Linguistics
   SPAN 414 – Spanish Phonology
   SPAN 440 – Teaching of Romance Languages
   SPAN 418 – The Evolution of Spanish

Publication:

Presentations:

Presentation given to students and faculty in Department of Spanish at the University of Florida, Gainesville.


Positions held:
2007-Present: Graduate assistant – Penn State Department of Spanish, Italian and Portuguese
2009-2012: Editorial assistant – Language Variation and Change (peer-reviewed journal)
Fall 2006: Instructor of Spanish – Penn State Department of Spanish, Italian and Portuguese