A Corpus-Driven Approach to the Infinitival Complementation Patterns of the Verb *HELP* in Recent Centuries

Jussi Ahola
University of Tampere
School of Language, Translation and Literary Studies
English Philology
Pro Gradu Thesis
May 2011

Tampereen yliopisto
Englantilainen filologia
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

AHOLA, JUSSI: A Corpus-Driven Approach to the Infinitival Complementation Patterns of the Verb *HELP* in Recent Centuries

Pro gradu –tutkielma, 91 sivua
Toukokuu 2011

Tämä pro gradu –tutkielma tarkastelee englannin kielen verbin *help* sekä sen taivutettujen muotojen *helps*, *helped* ja *helping* infinitiivikomplementteja brittienglannissa 1700-luvun alusta vuosituhannen vaihteeseen saakka. Tutkimuksen tavoitteina oli selvittää eri infinitiivikomplementtien esiintymistä eri aikakausina, löytää mahdollisia yhteyksiä komplementtien ja niiden välittämien semanttisten merkitysten välillä, selvittää *help*–verbin väitettyä kieliopillistumista sekä tutkia eri syntaktisten tekijöiden, kuten *horror aequi* –periaatteen, vaikutusta komplementin valintaan.

Tutkimuksen aineisto kerättiin kahdesta eri lähteestä: historiallinen aineisto kerättiin *Corpus of Late Modern English Texts* –korpuksesta (CLMET), joka kattaa vuodet 1710–1920, ja nykykielinen aineisto *British National Corpus* –korpuksesta (BNC). CLMET on jaettu kolmeen osaan, joista jokainen kattaa 70 vuoden ajanjakson. Näitä osia tarkasteltiin omina kokonaisuuksinaan, joiden välillä diakronisia vertailuja oli mahdollista tehdä. CLMET on rajoittunut lajityypiltään kaunokirjallisuuteen, mikä vaikutti rajoittavasti myös BNC:n käyttöön; ainoastaan BNC:n kaunokirjallinen materiaali hyväksyttiin aineistoksi. Tällä rajauksella varmistettiin eri aineistojen vertailtavuus.

Tutkielman sisältö on jaettu karkeasti kahteen osaan: ensimmäisen osan muodostavat neljä ensimmäistä kappaletta, jotka rakentavat tutkimuksen teoreettisen viitekehyksen. Nämä kappaleet käsittelevät kielitieteellistä metateoriaa, korpuslingvistiikkaa, sekä *help*-verbin infinitiivikomplementteihin liittyvää aiempaa tutkimusta. Seuraavat kaksi kappaletta muodostavat tutkielman analyysiosan. Aluksi aineisto käydään yksityiskohtaisesti läpi edeten kronologisessa järjestyksessä vanhimmasta uusimpaan. Tämän jälkeen esitetään vertailuja eri aikakausien välillä sekä esitellään tutkimuksen tulokset.

Tutkimuksella osoitetaan, että *help*-verbin infinitiivikomplementtien käytössä on tapahtunut huomattavia muutoksia tutkitulla ajanjaksolla. Yleisesti ottaen *help*-verbiä täydentävät infinitiivikomplementit ovat selkeästi yleistyneet, mikä on osittain selitettävissä *help*-verbin kieliopillistumisella. 1900-luvulla *to*-infinitiivien käyttö väheni, mutta puhtaiden infinitiivien (bare infinitive) käyttö yleistyi huomattavasti. *Horror aequi* –periaatteen vaikutus infinitiivikomplementtien valintaan osoitettiin olevan hyvin merkittävä.

Avainsanat: help, verbi, komplementaatio, infinitiivikomplementti, korpus.

Table of Contents

1. Introduction

> This verb is a corpus linguist's delight because its distribution in texts is so clearly influenced by stylistic, contextual, semantic and structural constraints, few of which are categorical in the sense that one variant form can be definitively excluded in a specific environment. (Mair 1995, 261)

The verb *HELP*[1]has lately received some attention from several linguists. A part of this attention can perhaps be attributed to its relatively high frequency of occurrence in the language; if we look at the word frequency list in the BNC, *HELP* is the 47th most frequent verb (lemmatized) in the whole corpus. For some linguists, interesting phenomena in the use of high-frequency items may be a more attractive subject of study than equally interesting things in the use of low-frequency, or even marginal items. In addition to being a relatively common verb in English, its syntactic features and their varying distribution in different environments are, perhaps, the main reason why it has been discussed by several scholars. The emergence of large and computerized corpora has aided the study of these features considerably and new corpus-based approaches have yielded interesting insights into the matter.

HELP can govern a variety of different complement patterns. The present study, however, is concerned with the infinitival complement patterns of the verb *HELP*. Consider sentences (1)–(4), all of which are authentic language samples taken from the British National Corpus:[2]

---

[1] For the sake of clarity, I use capital letters to distinguish the lemma of *HELP* from the word form *help* throughout this thesis. To follow standard conventions, however, direct quotations from other authors are reprinted as in the original regardless whether the word denotes the lemma or the word form.

[2] References to authentic language samples from the British National Corpus indicate the filename and the sentence unit of the sample in the corpus. Example sentence (1), for instance, can be found in the file A07 in the sentence unit number 1484. This convention is used with examples from the British National Corpus throughout this thesis.

(1) It has been argued that one could help to break that circle by integrating the school system. (A07 1484)
(2) The Zuckerman books help one to imagine how Roth has faced the reproach . . . (A05 1241)
(3) . . . coated with Scotchguard to help prevent the fabric from absorbing water. (CHK 683)
(4) If you want to help Mark reach his sponsorship target then phone him with your pledge on 0962 733987. (C88 1908)

The examples above serve to illustrate the different infinitival complement structures that the verb *HELP* can govern. These are the following:

HELP + *to*-infinitive clause
HELP + NP + *to*-infinitive clause
HELP + bare infinitive clause
HELP + NP + bare infinitive clause

This thesis aims to study various factors and constraints that condition the choice between these structures. This will be done with reference to the British variety of the language and to a timeframe spanning from the early 18th century to contemporary language. Changes in the distribution of complementation patterns within this timeframe and possible reasons for these changes will be discussed. As regards genre, the authentic language data for this thesis is collected from the domain of written prose; a suitable set of diachronic corpora from this genre was readily available for the present author. The thesis can be divided roughly into two main parts: the first part concentrates on providing theoretical background for the subject under study; the second part consists of the analysis of authentic corpus data.

As was noted above, *HELP* has been discussed somewhat extensively by several linguists; hence one of my goals is to scrutinize some claims made about its complementation in earlier literature. Some of the earlier literature has been based on corpora that can be regarded as rather small by modern standards; the emergence of large

and annotated electronic corpora can give new insights into the matter and allow further reaching statistical generalizations to be made. Furthermore, some new historical corpora have become available, which makes it possible to investigate trends in long-term diachronic development. With the help of these corpora, I can add a longer time perspective to the discussion which has largely been based on the Brown family of corpora representing language in the 1960s and 1990s.[3] This thesis does not discuss all the complements that *HELP* can govern, but is limited to a specific subset of these complements, i.e. the infinitival complements. This limitation of the topic stems from the fact that this approach prevails in recent literature discussing *HELP*. To be more specific, in this thesis I aim to:

i) Chart changes in the distribution of the infinitival complement patterns of the verb *HELP* from the early 18th century to present-day English.

ii) Examine whether different patterns convey different senses of the verb.

iii) Address the question whether *HELP* is undergoing a process of grammaticalization.

iv) Discuss different contextual factors that may condition the choice between competing complements. These factors may be syntactic, semantic, or extra-linguistic. In the data analysis special emphasis will be on possible effects caused by the phenomenon known as *horror aequi*.

By suggesting some answers to these research questions based on evidence from large present-day and historical corpora, I believe that I can add new and valuable perspectives to earlier treatments of this verb.

[3] The Brown family of corpora consists of four one-million-word matching written text corpora. LOB: BrE 1961; Brown: AmE 1961; F-LOB: BrE 1991; Frown: AmE 1991.

2. Methodology: corpora and corpus linguistics

This chapter presents the methodology on which the research part of this study is based. First, I will motivate the use of corpus data in linguistic research. Secondly, the concepts of *corpus* and *corpus linguistics* will be introduced and the two corpora used in this study presented. Finally, I will discuss the search heuristics that were used for gathering data and present some statistical measures used in this thesis.

2.1. Linguistic metatheory: why use corpora?

Broadly speaking there are two approaches to the scientific study of language: the theoretical approach and the descriptive approach (Hoffmann et al. 2008, 19). Both have their own traditions as concerns the focus of research and the kind of data that is favoured. As a way of motivating the type of data and methodology used in the present study, I will first introduce the theoretical approach and note some problems especially in the type of data used in this school of thought and then turn to the data used in the descriptive approach, under which the present study falls.

In a seminal article on linguistic metatheory Leech (1968, 88) identified three sources of data that can be used as a basis of linguistic research. These are: (a) corpora of authentic language; (b) elicited reactions of native speakers of the language; (c) introspection of the analyst who is a native speaker of the language (ibid.). At the time of the publication of the article the most influential approach to linguistics was, perhaps, that adopted in the school of transformational-generative, or theoretical, linguistics advanced by Noam Chomsky. The proponents of theoretical linguistics of the time were

mainly interested in the postulated and abstract linguistic competence, i.e. knowledge of a language in the brain, as opposed to linguistic performance realized in actual language use. Chomsky (1965, 10) advocated the use of introspection of a native speaker as a source of data because "limitations on performance imposed by organization of memory and bounds on memory" intervene between a person's linguistic competence and actual performance. As Leech (1968, 89) notes, these limitations result in some "nuisance variables, such as the occurrence of false starts, anacolutha, and hesitations," which in Chomsky's view restrict us from drawing linguistic generalizations based on corpus data. As Leech (1968) points out several times in his article, problems in the use of introspective data are plentiful. Since in most cases the informant and the analyst are the same person, perhaps the most problematic issue in the use of this type of data is "its unashamed subjectivism" (ibid., 87). Furthermore, two native speakers of a language may, and indeed in many instances will, disagree whether a given sentence is acceptable or not; hence they "have no common ground to investigate" (ibid., 89). Because of the subjectiveness of introspective data a fundamental requirement in science, replicability, is practically impossible to obtain (Hoffmann et al. 2008, 19); how to replicate the process and tests that have been based entirely on primary data of the introspective kind? Sinclair (1991, 6) commented sharply on these problems by stating that "[o]ne does not study all of botany by making artificial flowers." Leech claims that transformational-generative linguists favoured introspective data not because of its unmatched validity, but rather because of "the impracticability of obtaining other, more objective evidence" (Leech 1968, 93). Indeed, as Sinclair (1991, 1) points out, linguistics was "[s]tarved of adequate

data." Some of the problems in data-collection have, perhaps, subsided with certain technological advancements.

Many of the problems in the use of introspective data presented above are not prominent in research based on corpora of authentic language: the researcher does not have a twofold role as both an informant and analyst, the data is objective in the sense that it has been collected according to some explicit extra-linguistic criteria, and anyone with access to the corpus can replicate any previous research based on it. This type of data is what descriptive linguistics tends to favour. The "nuisance variables" in authentic data mentioned by Leech (1968, 89) can be placed under systematic research with the help of a suitable corpus and in many cases, in fact, it has turned out that their occurrence is by no means haphazard (ibid., 93).

Whereas theoretical linguistics is concerned with the postulated innate linguistic competence, descriptive linguistics focuses on actual linguistic performance. The difference can be described as the distinction between, in Saussure's terminology, *langue* and *parole*, or, in Chomsky's terminology, *competence* and *performance* (Sinclair 1991, 103). The fundamental difference between these approaches to language is eloquently worded by Hunston and Francis (2000, 17):

> . . . language is not a system that is realised in actual instances, but a set of actual instances that may be regarded as construing an approximate and ever-changing system.

The view of language quoted above serves as the foundation for the descriptive nature of this piece of research. Corpus data is, then, a part of that "set of actual instances" (ibid.) that constitutes the ever-changing and infinite totality of a language; it is used to describe language as it is actually used. It is precisely this descriptive nature of corpus data why it is given prominence in the present study as well.

2.2. Preliminaries: on corpora and corpus linguistics

Having motivated the use of corpus data in linguistic research in general and in the present study in particular, I will now turn to the nature of this data and the methodology surrounding it in more detail.

The word *corpus* was adopted into English from Latin and its original meaning was 'body of a person or animal' (*OED*, s.v. *corpus*, n. sense 1). Later it became to be used when referring to the complete works of an author and, during the 20th century, it was adopted into the terminology of linguistic research as meaning a 'collection of written and/or spoken language' (*OED*, s.v. *corpus*, n. sense 3.b). Sinclair (1996) offers the following definition of its meaning in modern linguistics:

> A *corpus* is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.

What is important in this definition is that the pieces of language that a corpus comprises of are explicitly selected. The criteria on which this selection is based are external to the texts themselves and decided by the persons compiling the corpus. These criteria are crafted to make the corpus representative of the language or language variety of which it aims to describe. As Hoffmann et al. (2008, 13) note,

> . . . although a corpus is only a (potentially quite small) subset of what it is supposed to represent, its function is to mirror the whole in such a way that linguists can use it to say something about the language variety that was sampled: observations on the basis of corpus data are generalized back to the whole from which the corpus was initially selected.

To compile a corpus that is truly representative of the language or variety that it aims to describe is a task that requires careful contemplation of what texts to include. Especially when compiling a general language corpus such as the British National Corpus (BNC),

for example, the decisions of what text types and how much of each of them to include in the corpus are extremely difficult to make and always subject to objections. It is arguably unlikely that two different groups compiling a general language corpus of present-day British English will arrive at exactly similar corpus make-ups. When the purpose of the corpus is to describe a language variety or text type that can be defined more precisely, this task becomes, perhaps, more straightforward. The Corpus of Late Modern English Texts (CLMET) aims to describe the language used in published British prose from the year 1710 to 1920; what to include in this corpus is not as big of a dilemma as in the case of the BNC, but still involves some problematic issues.

Once the corpus has been compiled and the criterion of representativeness met, researchers can observe the language in it and, not only describe the language in the corpus, but also extrapolate their findings into the language or variety that the corpus is a subset of (Hoffmann et al. 2008, 14). A truly representative corpus, then, allows researchers to make generalisations of the type "pattern A is more frequent than pattern B in language variety X" instead of restricting their statements merely to the linguistic material in the corpus.

Research based on the type of corpora presented above has become to be known as *corpus linguistics*. The term is generally understood to refer to a set of methods or a methodology rather than to a field of linguistics of its own (Lindquist 2009, 1). This methodology can be used to study different fields of linguistics, such as morphology and phonology, for example (Hoffmann et al. 2008, 19). Corpus linguistics is defined as "a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and

calculates" (Hunston and Francis 2000, 15). The emphasis on the use of automated data processing seems to prevail in definitions of *corpus linguistics* in the literature: "[c]orpus linguistics is the systematic study of linguistic phenomena using (machine-readable) collections of authentic language use, i.e. corpora" (Hoffmann et al. 2008, 18). It should be noted, however, that corpora can be said to have existed before the computer. These collections of texts, then, were in paper format, which, unarguably, posed certain restrictions on their use. Kennedy (1998, 13) has termed these as "pre-electronic corpora." As Lindqvist (2009, 3) notes, however, "nowadays *corpus* is almost always synonymous with *electronic corpus*. . . ." The emergence of large and annotated electronic corpora has certainly facilitated research based on large bodies of authentic language: today several corpora can be searched with few clicks of a mouse, whereas some decades ago huge amounts of text had to be gone through manually to find a fair number of instances of a given type.

2.3. The nature of the beast: restrictions in the use of corpora

Any corpus is ultimately a sample of a language or a language variety, i.e. a subset of all the data there is. The totality of the English language is infinite in nature; even if we could account for the whole lexis and all the possible combinations that words can appear in at a given point in time, every sentence can be, theoretically speaking, continued by using the coordinating conjunction *and*. This means that no matter what sort of innovations the future of data-collection, storage and retrieval techniques may hold, there

can never be a full corpus of English. This inevitable restriction of corpora should not, however, be considered as undermining the importance of corpus data:

> . . . whereas any corpus, however large, contains only an inconsiderable subset of the set of possible sentences of a language, this does not in any way diminish its importance as a tool of empirical confirmation: complete verifiability has long been acknowledged to be too high a goal in the testing of scientific theories. (Leech 1968, 94)

The evolution of corpus linguistics has witnessed an exponential rise in the size of corpora. The first computerized corpus, the Brown corpus, was built in the early 1960s and consists of one million words of British English (Leech et al. 2009, 24). Today, the largest electronic corpora are several hundreds of millions of words in size. Why, one might ask, this trend towards larger corpora?

The beauty of large corpora lies precisely in their size; the bigger the corpus, the better it will correspond to the language or variety of which it is a sample of. Though the full corpus of English is unachievable, a larger set of data is statistically more likely of containing more examples of a given phenomenon. This is important especially when one is working with low-frequency items. Thus when working with large and annotated electronic corpora, one does not wish, ideally, to take a sample of the corpus; this is taking a sample of something that is already a sample. The smaller the sample size gets, the less reliable extrapolations mirroring the situation in the entirety of that language or variety one can make. This applies, of course, only to quantitative research. As Hoffmann et al (2008, 18) state,

> [corpus linguistics] is an essentially quantitative method, meaning that corpus linguists tend to count features of language (or have the computer count them) as part of their analysis of linguistic features.

The use of large corpora instead of smaller samples can, of course, get rather wearisome if desired data from the corpora cannot be extracted automatically with the help of computer software with high recall and precision.[4] It is not desirable to weed out irrelevant tokens out of tens of thousands of hits that were produced by bad precision.

Quantitative work is, however, just the first part of the analysis. Though technological advances in the area of corpus linguistics have been relatively rapid, even modern corpora cannot provide detailed information on semantic issues, for example. To be able to say something about the nature of the data, one needs to employ qualitative methods. For this work, the trained eye of a linguist cannot be matched by computer software. As stated by Sinclair, (1991, 39) introspection is better used "in evaluating evidence rather than creating it." When the number of collected tokens is unmanageable for manual inspection, a smaller subset of the collected tokens is a better approach for the qualitative work. Once the statistics of the distribution of certain patterns, say *to*-infinitives and bare infinitives after *HELP*, have been put together, the data should be investigated more closely to find out differences in meaning between the patterns, for example (Hoffmann et al 2008, 18). In this study, I will employ this type of mixture of quantitative and qualitative methods.

2.4. Corpora used for the present study

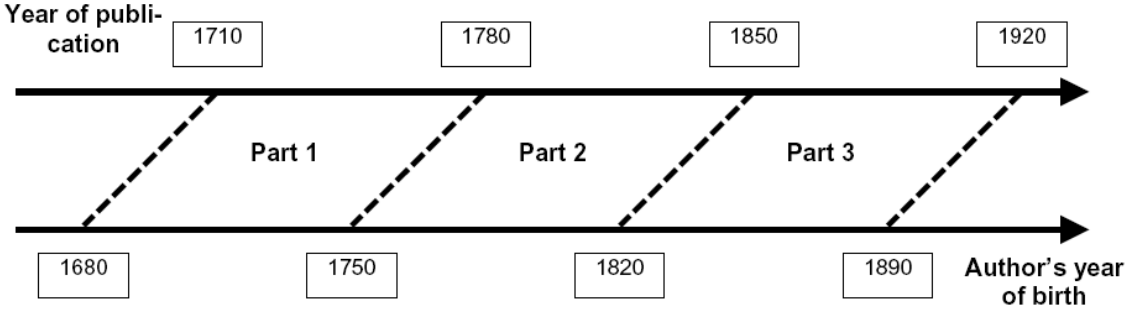
This section introduces the two corpora used in this study. These are the Corpus of Late Modern English Texts and the British National Corpus.

[4] The concepts of *recall* and *precision* will be explained below in Section 2.5.1.

2.4.1. The Corpus of Late Modern English Texts

The Corpus of Late Modern English Texts (abbreviated as CLMET) provides the basis for the historical part of the diachronic analysis in the present study. CLMET consists of approximately 10 million words of mostly literary texts published in Britain between the years 1710 to 1920 (De Smet 2005, 70). The texts in the corpus were collected from online resources *Project Gutenberg* and the *Oxford Text Archive*. The corpus is further divided into three sub-sections each covering a time-span of 70 years (ibid.). As was noted above, corpora should be compiled with some language-external and specific principles in mind. In the case of the CLMET, these criteria were, as described in De Smet (2005, 70–72), the following: (a) texts within each 70-year-sub-section were written by authors born within a corresponding time-frame preceding the sub-section by 30 years; this is illustrated in Figure 1 below; (b) all accepted authors were British and native English speakers; (c) a maximum of 200,000 words was accepted from each author; (d) whenever possible, non-literary texts from lower registers were favoured over literary texts from higher registers; furthermore, some attention was paid to include texts by women authors.

Figure 1. Matching the authors' date of birth with the date of published material in the CLMET, adopted directly from De Smet (2005, 71).



De Smet (2005, 72) notes that regardless of the measures described above, the CLMET "continues to be biased to literary texts written by higher class male adults."

As concerns the representativity of the CLMET, the present study considers it to be representative of published British English from 1710 to 1920 with a bias to literary texts written by upper class male adults. When data from the CLMET is contrasted to data representing present-day English, the text type that the CLMET represents will be taken into account.

Corpora are, by definition, merely large bodies of text and cannot be easily worked with as such; one needs special software to be able to conveniently extract desired data from a corpus. In this study the CLMET was searched using a program called *MonoConc*, which extracts data from the corpus into plain text files. These files were then sorted and analyzed manually.

2.4.2. The British National Corpus

The British National Corpus (abbreviated as BNC) is perhaps the most famous one among the "second generation mega-corpora" (Kennedy 1998, 45). It was created to be representative of British English of the late 20th century (Hoffmann et al. 2008, 27) and contains slightly under 100 million words of written and spoken language. Unlike the CLMET, the BNC was created as a general language corpus, which means that a lot of planning went into deciding what text types and how much of each of them should go into the corpus; a more detailed description of the included text types can be found in Burnard (2007). If we hold the amount of language that is produced as the most important

criterion, it is clear that the vast majority of text in any general language corpus should ideally be spontaneous spoken language; most people interact considerably more through spontaneous conversations than through producing planned written text (Hoffmann et al. 2008, 16). Unfortunately the collection and transcription of spoken language is considerably more time-consuming and expensive than in the case of written language. Because of limited resources, the compilers of the BNC decided that about 10% of the material in the corpus should consist of spoken texts while 90% was left for written material (Burnard 2007). The corpus is further divided into specific sub-sections according to text types; this means that different parts of the corpus can be studied separately or contrasted to each other.

The BNC is a grammatically annotated corpus. This means that in addition to the actual primary material in the corpus, all words have been annotated by adding metadata about their grammatical class into the corpus. Furthermore, the metadata includes information on the headword, i.e. all the inflected forms of a given word (*help*, *helps*, *helped* and *helping* are listed under the headword *HELP*, for example). In the case of the current XML Edition of the BNC, this information is encoded into the corpus as "tags" using a markup language called *XML*[5] (Burnard 2007). With the help of this information the BNC allows searches for specific parts of speech, headwords, or the combination of both, i.e. lemma searches (Hoffmann et al. 2008, 38–41). However, care should be taken whenever these tags are used in corpus queries; the automatic annotation for parts of speech is not perfect and some errors in the corpus are inevitable (Burnard 2007).

Because of the metatextual information in the BNC it is not, one could argue, merely a large body of text. The XML format of the corpus is, however, practically

[5] XML is shorthand for *Extensible Markup Language*.

unreadable; software for viewing and extracting data from the corpus is hence needed. In the present study the BNC was searched using an interface called *BNCweb*. Searches that utilize the grammatical annotation can be made with the interface and some useful post-query options are available. Two query languages are available in *BNCweb*: the Simple Query Syntax and the CQP Query Syntax. The former is relatively intuitive and covers most query scenarios; the latter is more complex, but it allows more powerful and elaborate queries to be performed. Both query languages were used in this study.

In this study, the searches in the BNC were restricted to a specific text type. This text type is the Imaginative prose, which was selected because it matches closely with the text type of the CLMET corpus.

2.5. On corpus search heuristics

This section introduces the concepts of *precision* and *recall* and presents the search strings that were used to extract data. Furthermore, in this section I will present the statistical measures used in this thesis.

2.5.1. Recall and precision

In a corpus linguist's dream, the software used for extracting desired data from a corpus would find all the tokens and only the tokens that are relevant for the research question under investigation (Hoffmann et al. 2008, 78). In reality, this is hardly ever so. Sometimes undesired tokens find their way into the query results and have to be removed

from the data; there are problems in the precision of the search. In other cases all the relevant instances are not collected by the used search heuristics; the recall of the search is not perfect.

In Ball (1994, 295), *recall* is defined as "the proportion of relevant information that was retrieved" and *precision* as "the proportion of retrieved material that is relevant." The precision of a given search is relatively easy to inspect: some of the concordances should be manually inspected to see whether there are any undesired tokens in the data. These can always be dealt with by removing them from the data. This, of course, can be rather time-consuming when working with large bodies of text and high-frequency items and is therefore not desirable. The level of recall is harder check; it is difficult to know what one is missing. As Ball (ibid.) notes, "it is generally impossible for the analyst to know what has been missed without analysing the entire corpus by hand." To avoid such an overwhelming task, researchers generally aim to improve their search strings by trying to account for all the relevant instances in their search strategy. This, however, is not a guarantee that instances of a given type will not be missed. Hence the crafting of good search strings is a matter of great importance in any corpus based study (Ball 1994, 296).

2.5.2. Search strings: finding an optimal balance between recall and precision

Especially with large corpora and high-frequency items, poor precision may lead to frustration in the process of manually weeding out the irrelevant tokens out of all the hits produced by a search. This may lead to the analyst narrowing down the search criteria, which, again, often decreases the recall of the search (Ball, 1994, 295). Generally

speaking, improving the recall of a search may decrease its precision and vice versa; finding an optimal balance between recall and precision is a question that always needs to be addressed when one is working out the search strings.

Within the confines of the present study, it has become evident that different corpora require different search heuristics, which, in turn, will result in altogether different levels of recall and precision. The most important factor in this is whether the corpus is annotated or not. The BNC is, for that matter, tagged for grammatical classes. This allows more elaborate searches to be performed with it. Patterns with *HELP* which do not include the intervening NP object are relatively straightforward to search for with the help of *BNCweb* interface to the BNC. The following search strings work in *BNCweb*'s Simple Query Syntax:

Table 1. Search strings for the infinitival complements of *HELP* without the object NP in the *BNCweb* Simple Query Syntax.

Pattern	Search string
HELP + *to*-infinitive	{*help*}_V* *to* _V*I
HELP + bare infinitive	{*help*}_V* _V*I

These search strings correspond to any form of the verb *HELP* + (*to*) + any verb in the infinitive form. The precision of the above search strings appears to be on a sound level; the manual inspection of the concordances revealed only a few tokens that are irrelevant for the present study.[6] As was noted above, the level of recall is harder to inspect; it is difficult to know what one is missing. Through deductive reasoning, however, we can expect the level of recall to be sound, too. What the search strings in Table 1 do not collect, however, are instances where insertions have occurred between the matrix verb

[6] These irrelevant tokens are passives, which, however, are relevant as complement patterns with the object NP, and extraposed infinitive clauses. The reason why extraposed infinitive clauses are not relevant for the present study will be discussed in due course.

and the infinitive clause. To tackle this problem, additional searches that allow one to three words to intervene in the pattern were carried out. This combination of searches should collect most of the instances of any form of the verb *HELP* followed by either a *to*-infinitive or a bare infinitive in the corpus.

The infinitival complement patterns that have the intervening object NP between the matrix verb and the verb in the lower clause are trickier to search for; one cannot easily allow the object NP to intervene in the pattern. This is possible, however, with *BNCweb* by specifying the NP slot as, say, a string of up to six words with none of the intervening words a verb and the last word a noun or a pronoun.[7] For queries of this type *BNCweb*'s CQP Query Syntax has to be employed. The CQP Query Syntax is a more powerful, yet less intuitive, query language than the Simple Query Language; it allows even more elaborate searches to be performed with it (Hoffmann et al. 2008, 215). Search strings with the object NP that were described above are translated into the CQP Query Syntax in Table 2 below:

Table 2. Search strings for the infinitival complements of *HELP* with the object NP in *BNCweb*'s CQP Query Syntax.[8]

Pattern	Search string
HELP + NP + *to*-infinitive	[hw="help"%c & pos="V.*"] [pos !="V.*"] {0,5} [pos="N.*|PN.*"] [word="to"%c] [pos="V.*I"]
HELP + NP + bare infinitive	[hw="help"%c & pos="V.*"] [pos !="V.*"]{0,5} [pos="N.*|PN.*"] [pos="V.*I"]

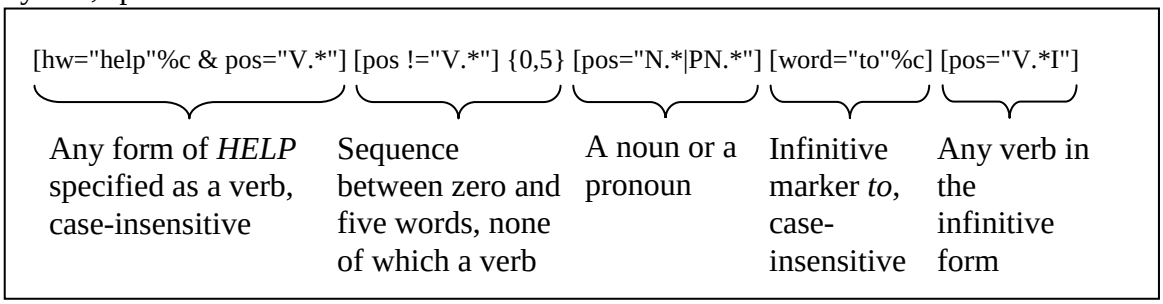
Though these search strings may appear somewhat cryptic for an unadjusted eye, they follow a specific logic and correspond to sequences of any form of the verb *HELP* + NP

[7] This method does not collect NPs that have a verb in them, for example as part of a non-finite clause or a relative clause. Such NPs, however, should be very infrequent in this slot.

[8] Hoffmann et al. 2008, especially Chapter 12, has been an invaluable guide when crafting these search strings.

+ (*to*) + any verb in the infinitive form. Figure 2 below will clarify this using the pattern with the *to*-infinitive:

Figure 2. The search string for the *HELP* + NP + *to*-infinitive pattern with CQP Query Syntax, spelled out in constituents.



The search string for the pattern with the bare infinitive is identical to the above with the exception that the infinitive marker *to* is omitted.

The precision of the search strings with the intervening NP object is, evidently, on a good level; going through the concordances, I have identified only few undesired tokens in the data.[9] Though the recall of the search strings presented above is, again, harder to inspect, it seems that they provide high recall at a minimal cost in terms of precision. This, of course, is essential for any automated text analysis. The search strings conveniently collect hundreds of tokens with the object NP, including NPs which are several words long:

(5) Egbert Rose was helping the sailor-suited youngster on his right to fortify his castle. (H8A 522)

Multi-word object NPs such as *the sailor-suited youngster on his right* might not have been easily collected without resorting to the search heuristics that have been discussed

[9] The search string for the NP + *to*-infinitive complements collects some undesired tokens of the type *HELP* + PP + *to*-infinitive, where the PP does not function as a direct object of *HELP* but is rather an adjunct that modifies the matrix verb. Other irrelevant tokens are due to extraposition and mistaggings. One irrelevant token collected by the search string for the NP + bare infinitive complements was mistagged: . . . *helping the cotton export to flow* . . . (FPF 753), where *export* is tagged as a verb; this token is, however, relevant as an NP + *to*-infinitive complement.

above. To collect passive transformations, an additional query *_VB* helped* was performed; this collects all instances of the verb *BE* followed by *helped*.

The CLMET is not annotated for grammatical classes, but rather is a plain text corpus. This means that a completely different method for collecting data has to be used; in the case of CLMET one can only search for concrete words or parts of words. The searches that were performed collected all the instances of the inflected forms of *HELP*: *help*, *helps*, *helped*, and *helping*. These yield results that are perfect in terms of recall, but not very accurate in terms of precision. The vast majority of the tokens collected by the searches are irrelevant for the present study. These include, for example, non-infinitival complement patterns and cases where *help* is a noun. Fortunately, problems with precision are not as critical as problems with recall; the undesired tokens can always be dealt with by manually removing them from the material. This can, however, be rather time-consuming especially with big corpora and high-frequency items. Crafting good search strings is always a question of trade-off between recall and precision.

2.5.3. Normalizing frequencies

As was noted above, corpus linguistics is essentially a quantitative method, meaning that corpus linguists often quantify their results and compare the numbers obtained from searches in different sections of a corpus or from altogether different corpora. One of the caveats in this method is to trust absolute numbers, or *raw frequencies*, as a basis of comparison. The different sections of a corpus or different corpora are often not of equal size; to compare raw frequencies in these cases is misleading (Biber et al. 1998, 265).

When comparing results, it is essential to take into to account the different sizes of specific parts of the corpus or different corpora (Hoffmann et al. 2008, 70–71). This can be done by choosing a specific number of words as a basis for the comparison and relativizing the raw frequencies to that number. This procedure results in *normalized frequencies* (NF). The arithmetic to arrive at normalized frequencies is presented below, with one million words chosen as the basis of comparison:

$$\text{frequency pmw} = \frac{\text{number of tokens}}{\text{number of words}} \text{ x } 1{,}000{,}000$$

Frequency per million words (pmw) is calculated by dividing the number of relevant tokens by the number of words in the corpus or a relevant part of the corpus and multiplying the result by 1,000,000.

I will use this method of normalizing frequencies in my analysis of the corpus data. The number of words used as the basis of comparison will be 1 million.

2.5.4. The chi-square test

In this thesis the chi-square test will be used to test whether a certain difference is statistically significant or not. The test determines the level of certainty that we can have in the observed difference being either statistically marked or merely a product of chance, i.e. caused by sampling variation (Hoffmann et al. 2008, 83–86). As has been emphasized, the corpus is nothing more than a sample. McEnery and Wilson (2001, 84–85) explain the test in the following manner:

> Very simply, the chi-squared test compares the difference between the actual frequencies which have been observed in the corpus (the *observed* frequencies) and those which one would

> expect if no factor other than chance had been operating to affect the frequencies (the *expected* frequencies). . . . [T]he greater the difference between the observed frequencies and the expected frequencies, the more likely it is that the observed frequencies are being influenced by something other than chance. . . .

To exemplify the matter, let us assume that we find 11 instances of construction A and 15 instances of construction B in environment X, and 23 instances of construction A and 12 instances of construction B in environment Y. From these numbers a 2 x 2 table can be formed:

Table 3. 2 x 2 table for the chi-square test.

Environment	A	B
X	11	15
Y	23	12

Intuitively there seems to be a difference in the distribution of A and B caused by the environment in which they occur. To test whether this difference is statistically significant or not, the chi-square test can be used.[10] The test gives an X^2 value of 2.43 for our example. Traditionally some critical values for the chi-square test are acknowledged:

Table 4. Critical values for the chi-square test (Hoffmann et al. 2008, 85)

Critical value	Confidence level	*p* value	Level of significance
3.841	95 % confidence	$p < .05$	minimum level of significance
6.635	99 % confidence	$p < .01$	high significance
10.827	99.9 % confidence	$p < .001$	very high significance

[10] For chi-square calculations, I will use an online resource called *Corpus Frequency Test Wizard*, which provides a neat and simple interface. The *Corpus Frequency Test Wizard* can be found at http://sigil.collocations.de/wizard.html. As described in Hoffmann et al. (2008, 85), the frequency comparison test "chooses between chi-squared (X^2) and log-likelihood (G^2), depending on which test is deemed more accurate for your data." In this thesis, the chi-square test was deemed more accurate for all calculations. In the *Corpus Frequency Test Wizard* numbers must be inserted as "positive" cases out of the whole sample. In this example, I treated instances of construction A as positive cases, and hence the numbers 11 and 23 go into the *Frequency count* column, and the numbers 26 and 35, which are the whole sample sizes, go into the *Sample size* column. Naturally, it does not make any difference whether A or B are treated as "positive" cases.

The X^2 value of 2.43 is not above the critical value of 3.841, and hence the result is not above the minimum level of significance at $p < .05$. In other words, our confidence in the observed difference being a result of the environment is less than 95%, which is traditionally considered the minimum level of significance.

3. Earlier literature: *HELP* in dictionaries and grammars

In this chapter I will introduce the way *HELP* is treated in the *Oxford English Dictionary* and some grammars. In the *OED* analysis special emphasis will be on the different senses and complementation patterns of *HELP*.

3.1. *HELP* in the *OED*

Dictionaries provide a solid grounding for the study of the meaning of words. Since the *Oxford English Dictionary* (*OED*) is oftentimes regarded as the most authoritative etymological dictionary of the English language, it was used as the main source of lexicographical information for the present study as well.

The fact that *HELP* is a common verb in English which has a long history is reflected in the relatively high number of senses and sub-senses that it has; the *OED* lists altogether 11 senses and numerous sub-senses under its entry for *HELP* as a verb. Sinclair has tentatively proposed that there is a relationship between the frequency of occurrence of a word and the number of different senses that the word has:

> Frequent words have, in general, a more complex set of senses than infrequent words. If we divide and number senses in the conventional dictionary manner, we may discover a statistical relationship between the number of occurrences of a word and the number of different senses it realizes. Hence, the accumulation of instances of a frequent word is not just more of the same, but ever more clear evidence of complexity. (Sinclair 1991, 101)

Since this study, however, concerns a specific subset of the complements that *HELP* can select, i.e. the infinitival complements, the main focus in the *OED* entry for *HELP* is on those senses which, according to the entry, can select these complementation patterns. As

it happens, the *OED* specifies its sense 5 (*OED* s.v. *help*, v. sense 5.) as selecting infinitival clauses. For this reason I decided to cite only the relevant part of the *OED* entry here. Sense 5 of the *OED* entry for *HELP* as a verb is given in Table 5 below. Sense 5.c. is deemed irrelevant for this study because its last occurrence is recorded in the 16th century (*OED* s.v. *help*, v. sense 5.c.). Hence it was not included in Table 5.

Table 5. Sense 5 of the *OED* entry for *HELP* (*OED* s.v. *help*, v. sense 5.)

Senses of *HELP*	Illustrations	Complement patterns
5. With *inf.* or *clause*:		
a. With *inf.* alone. (This may either arise through ellipsis of the object in b, or may be a use of sense 4 with inf. obj.) In this and 5b the infinitive has normally *to*, which however from 16th c. is often omitted: this is now a common *colloq.* form.	**1.** He help'd to bury whom he help'd to starve. (1735, Pope *Epist. to Arbuthnot*)	*to*-infinitive
	2. All the leaves that helped nourish it. (1853, T. T. Lynch *Lect. Self Improvem.*)	bare infinitive
b. With *obj.* and *inf.* To aid or assist (a person *to do* something). (See sense 1.)	**3.** The hereditary enemies of his house had helped him to mount a throne. (1855, T. B. Macaulay *Hist. Eng.*)	NP + *to*-infinitive
	4. I would fain stay and help thee tend him. (1852, M. Arnold *Empedocles on Etna*)	NP + bare infinitive

The examples in the *Illustrations* column were chosen to exemplify the different patterns that were found among all the examples given in the *OED*. Since the *OED* does not mark different complement patters explicitly using Phrase Structure labels, these were worked out from the examples in the middle column and are presented in the *Complement patterns* column.

The *OED* senses 5.a and 5.b. are illustrated with examples of both *to*-infinitive and bare infinitive constructions. According to the *OED* (s.v. *help*, v. 5.a), "the infinitive has

normally *to*, which however from 16th c. is often omitted: this is now a common *colloq.* form." In other words, the *OED* treats the *to*-infinitive as the original form, while the bare infinitive has developed later and is a colloquial form. Besides this subtle stylistic difference, the *OED* does not spell out any semantic differences between the two types of infinitives.

As regards the different senses and their compatibility with different complement patterns, the patterns with the infinitive only (*OED* s.v. *help* v. sense 5.a) may be the result of the omission of the object NP in sense 5.b, which can be characterized as the prototypical meaning of *HELP*, or may be a case of sense 4 with an infinitive object. Sense 4 is described as "[t]o make (an action, process, condition, etc.) more effectual; to assist in bringing about; to further, promote" (*OED* s.v. *help* v. sense 4). Forms that have the object NP (*OED* sense 5.b) manifest the prototypical meaning of *help*. Leech et al. (2009, 190) describe this prototypical meaning as "the core trivalent meaning of the verb *help* ('A helps B to do C')."

According to the Bolinger's Principle, we would expect that "a difference in syntactic form always spells a difference in meaning" (Bolinger 1968, 127). When the different infinitival complement patterns of *HELP* are matched with different senses on the basis of the *OED* analysis, the following picture emerges:

Table 6. Matching the *OED* senses with the infinitival complement patterns.

Sense	Complements
To aid or assist (a person to do something)	*to*-infinitive NP + *to*-infinitive bare infinitive NP + bare infinitive
To make (an action, process, condition, etc.) more effectual; to assist in bringing about; to further, promote.	*to*-infinitive bare infinitive

According to the *OED*, the first sense, 'to aid or assist,' can select as its complements all four of the possible infinitival complements. The second sense, which has been characterized by Mair (2002, 124) as "'contribute to/provide a favourable environment for,'" is, according to the *OED*, more likely to select complements without the object NP. This sense has been referred to as the more grammaticalized one (Leech et al. 2009, 190–191), which is highlighted by the fact that it is realized in patterns without the object NP that is present in the core meaning 'A helps B to do C.' For ease of reference, I will refer to the prototypical sense as the 'aid' sense, and the more grammaticalized sense as the 'contribute to' sense.

3.2. *HELP* in grammars: syntactic characteristics

Biber et al. (1999, 694) recognise five major infinitival patterns in English. As regards *HELP*, they state that "[t]he verb *help* is particularly noteworthy in that it can occur with

four of the five possible structural patterns"[11] (Biber et al. 1999, 708). As has been already noted above, these infinitival complementation patterns are the following:

HELP + *to*-infinitive clause
HELP + NP + *to*-infinitive clause
HELP + bare infinitive clause
HELP + NP + bare infinitive clause

In other words, *HELP* can govern either a full infinitive (*to*-infinitive) or a bare infinitive and can occur either with or without an intervening NP between the matrix verb and the infinitival clause. In terms of semantics, Biber et al. (1999, 693) place *HELP* in the category of "verbs of modality or causation." Furthermore, it is noted that *HELP* can control an extraposed *to*-infinitive clause in post-predicate position (Biber et al. 1999, 714). Consider example (6), taken from Biber et al. (ibid.):

(6) It may also help to set out some educational objectives against which the ideas will be tested.

In the sentence above the infinitival clause *to set out some educational objectives against which the ideas will be tested* has been extracted from its subject position in the higher clause to post-predicate position; to use terminology from the field of transformational linguistics, a transformation has occurred. A pleonastic *it* fills the subject position and it functions as the structural subject. The infinitival clause in the lower clause still functions as the logical subject. A canonical version of (6) would hence be (7):

(7) To set out some educational objectives against which the ideas will be tested may also help.

In the canonical version *HELP* is intransitive; it controls a zero-complement. Biber et al. note, however, that infinitival clauses are much more common as extraposed than in pre-

[11] The fifth major infinitival complementation pattern recognized by Biber et al. (1999, 694) is verb + *for* NP + *to*-clause, with which *HELP* does not occur. Verbs such as *WAIT* and *ASK*, for example, can select this pattern.

predicate position and recommend that "[e]xtraposed constructions should be regarded as the unmarked choice whenever a *to*-clause functions as logical subject of the main clause" (Biber et al. 1999, 725). This can, perhaps, be partly accounted by the principle of end-weight in the English language, according to which heavy elements tend to be placed towards the end of a sentence (Leech and Svartvik 2002, 210–211). In most cases when a *to*-infinitive clause does occur in pre-predicate position, it either "marks the information in the *to*-clause as being topical and is anaphoric to the preceding discourse," or is used because there already is a complex construction in the lower clause (Biber et al. 1999, 725–727).

In environments where an extraposition has occurred, such as (6) above, the surface structure suggests that *HELP* is complemented by a *to*-infinitive clause. A closer look into the argument structure of *HELP*, however, reveals that this is not necessarily the case. Haegeman (1991, 51–56) has argued that expletives *it* and *there* are semantically empty slot-fillers and hence should not be considered as part of the argument structure of the predicate; the predicate does not give them a theta role. This puts example (6) under new light: if the expletive *it* is not an argument of *HELP*, its only argument is the infinitive clause *to set out some educational objectives against which the ideas will be tested*, which would then be the subject argument; traditionally subject arguments are not considered complements. Hence we can conclude that in (6) *HELP* is, on a closer look, intransitive and controls a zero complement. If extraposed sentences with *HELP* occur in the corpus data, I will treat them as intransitive and consequently deem them irrelevant for the purposes of this study. However, in environments where *it* is referential, i.e. anaphoric with preceding discourse, a following infinitive is normally a complement.

Quirk et al. (1985) do not have a separate discussion on verb + bare infinitive complements, but recognise *HELP* in all other relevant verb complementation types (Quirk et al. 1985, 1187–1206). Furthermore, it is noted that the passive transformation can be applied to complements with the object NP (ibid., 1176).[12] When taken strictly, the following statement made by Quirk et al. (ibid., 1187), however, does not seem to hold water: "When a subjectless infinitive clause is direct object, the 'understood' subject of the infinitive clause is always the same as the subject of the superordinate clause." Consider sentence (8) below, taken from the BNC:

(8) My mother helped [PRO] to cater for the funeral tea, which were community affairs with everyone helping. (BN6 1505)

Surely the understood subject of the infinitive clause, which is marked with PRO in (8), is not co-referential with subject of the higher clause, i.e. *My mother*. The mother is not helping herself, but rather someone or something else, the grieving family for instance, at the funeral reception. In other words, *HELP* in (8) does not exhibit subject control. The term suggested in the literature for this type of control is NP_1 *plus* control, where the PRO is partly controlled by an "unexpressed but understood NP" (Rudanko 1996, 34).[13]

Leech and Svartvik (2002, 406–408) comment on the syntactic properties of *HELP* by stating that *HELP* selects as its complements infinitive clauses and object + infinitive clause constructions and continue that "[*h*]*elp* can be used with a to-infinitive or a bare infinitive" (ibid., 406).

[12] In this study, passives are related to their active counterparts when the study of argument structure and complementation is in question; this approach is motivated in Perlmutter and Soames (1979, 30-33).
[13] The notion of control is not discussed in any detail in this thesis. I would guide my reader to Davies and Dubinsky (2004) for a more extensive treatment.

4. Earlier literature: factors bearing on the complementation of *HELP*

Because of its distinctive syntactic characteristics, the verb *HELP* has received a fair amount of attention in the literature. An examination of the literature reveals that this discussion has been mostly about the infinitival complement patterns of the verb. As noted by Mair, (1995, 261) the distribution of the complements is "clearly influenced by stylistic, contextual, semantic and structural constraints. . . ." In the following sections I will present some claims made about these constraints in earlier literature.

4.1. Present-day synchronic variation based on language variety

Since *HELP* can govern either a *to*-infinitive or a bare infinitive, some studies have been conducted to find out which of the two infinitive constructions is more prominent in different varieties of present-day English, most notably in British English and American English.

McEnery and Xiao (2005, 163–166) study the impact of language variety by comparing usage between three British and three American corpora representing language in the latter part of the 20th century. The discussion of present-day written English is based on data from the FLOB and Frown corpora, which represent language in 1991 in Britain and the U.S., respectively. According to their findings, the bare infinitive accounted for 60% of all infinitival complements after *HELP* in written British English in 1991; the corresponding figure for American English was 82%. In spoken language the percentages were roughly similar. The fact that the patterns without *to* are the statistically

favoured option in both BrE and AmE is confirmed by Mair (2002, 121–126). Mair's discussion of written language is based on the FLOB and Frown corpora while the spoken data comes from the spoken-demographic section of the BNC. These findings are repeated in a more recent treatment of the verb *HELP* by Leech et al. (2009, 187–193). Leech et al. (2009, 189) quite correctly raise the question "whether a different degree of preference for the same option should be considered a salient regional contrast." Indeed, in the case of infinitival complements after *HELP* in present-day English, the contrast seems not to be between preferences for different constructions, but rather between different degrees of preference for the same construction.

The American preference for the bare infinitive is noted by some grammar books, too. According to Quirk et al. (1985, 1206), "[o]f the two constructions with *help*, that with *to* is more common in BrE, and that without *to* is more common in AmE." According to the data from the early 1990s presented by other scholars discussed earlier in this section, however, this statement is not entirely accurate anymore: the bare infinitive is nowadays the favoured option in BrE as well. The publication date of the grammar by Quirk et al. (ibid.) has to be taken into account, of course; they were presumably describing the situation based on data from earlier decades. Biber et al. (1999, 735) are more accurate by stating that the bare infinitive is more common than the *to*-infinitive in both varieties, but the preference for the bare infinitive is remarkably strong in AmE.

4.2. Diachronic development in recent decades

The prevailing view in the recent literature is that language change has had a significant effect on the language user's choice between the two types of infinitive after *HELP*. According to all the studies consulted that address the question, the fact that the bare infinitive is the more favoured option across varieties today seems to be the result of a relatively recent development. According to findings by Mair (2002, 122), which are based on matching corpora of written text from the 1960s and the 1990s,[14] the *to*-infinitive was the statistical norm in BrE in the 1960s, but by the 1990s the bare infinitive had gained ground and was the more favoured option. In AmE, the bare infinitive was the norm already in the 1960s and became proportionally even more used by the 1990s. Table 7 below sums up Mair's (ibid.) findings:

Table 7. Mair's (ibid.) findings on diachronic change in recent decades

Data	*To*-infinitive as % of total	Bare infinitive as % of total
BrE 1961	78 %	22 %
BrE 1991	39 %	61 %
AmE 1961	31 %	69 %
AmE 1991	18 %	82 %

These findings are replicated in McEnery and Xiao (2005, 166–168). Leech et al. (2009, 189) add data from BrE in 1931 to the discussion, according to which the bare infinitive was even less frequent in the first half of the 20th century than it was in the second half. Mair (2002, 123–124) also studied long-term diachronic development with the help of the *OED* citation database. The bare infinitive seems to have been only marginally used from

[14] The LOB and F-LOB for BrE in 1961 and 1991, respectively, and the Brown and Frown for AmE in 1961 and 1991, respectively.

the early 17th century until the dawn of the 20th century, after which its proportion of all the infinitival complements started rising, surpassing the *to*-infinitive at some point during the second half of the century (ibid.).

According to the findings presented above, there was a parallel diachronic development between the 1960s and 1990s which favoured the bare infinitive over the *to*-infinitive after *HELP*. In BrE the change was statistically clearly more significant than in AmE because "by 1961 AmE was already much more tolerant of bare infinitives than BrE" (McEnery and Xiao 2005, 167). This starting point paved the way for a more marked change in BrE during the next 30 years. The present-day statistical preference for the same option is, according to these earlier studies, a result of changes in the complementation system after the 1960s.

4.3. Grammaticalization

The last few sections have established the point that the bare infinitive is nowadays the proportionally more frequently used of the two infinitive constructions in both BrE and AmE. This, however, is not all that there is to the matter. The steep overall increase in the number of the bare infinitive constructions after *HELP* cannot be accounted by the small decrease in the number of the *to*-infinitive constructions; as Leech et al (2009, 189) note, "the variation observed between the two types of infinitive is not a zero-sum game, in which one variant gains to the extent that the other one loses." There has been, indeed, a significant overall increase in the total number of infinitives after *HELP*; the frequency of the *to*-infinitives has decreased slightly, but the total number of infinitival complements

after *HELP* has increased, which is due to the dramatic increase in the number of the bare infinitive constructions (ibid.). Mair (2002, 123–126) explains this increase by claiming that *HELP* is undergoing a process of grammaticalization. Mair (ibid., 123–124) states that ". . . the meaning of *help* has broadened, from 'somebody lends support to somebody else in performing some task' to a more general notion of 'contribute to/provide a favourable environment for'", the latter meaning being "so general and abstract that it approaches those typically associated with grammatical categories." Mair's statement about the semantic bleaching is in line with Sinclair's tentative proposal, according to which the increasing frequency of a word might lead to its delexicalization:

> There is a broad general tendency for frequent words, or frequent senses of words, to have less of a clear and independent meaning than less frequent words or senses. These meanings of frequent words are difficult to identify and explain; and, with the very frequent words, we are reduced to talking about uses rather than meanings. The tendency can be seen as a progressive delexicalization, or reduction of the distinctive contribution made by that word to the meaning. (Sinclair 1991, 113)

In other words, the fact that *HELP* is a frequent word in English is contributing to the grammaticalization of its meaning.

The fact that the bare infinitive, which is typical after modals in English, is nowadays the favoured option after *HELP* can be treated as evidence of its syntactic grammaticalization (Mair 2002, 124–125). In a recent treatment Leech et al. (2009, 190–193) discuss grammaticalization with respect to differences between written and spoken language. Firstly, they note that the relative frequencies of infinitives after *HELP* in general are lower in spoken than in written language (Leech et al. 2009, 191). Their second point is connected to the object NP. Because the core meaning of *HELP* has three parts, as in "'A helps B to do C'", Leech et al. treat forms without the object NP between

the matrix and the infinitive clause as more likely of manifesting the more grammaticalized 'contribute to' sense of the verb[15] (Leech et al. 2009, 190). Backing up their claims with evidence from corpora, Leech et al. (ibid., 190–191) state that in spoken language forms with the object NP are more frequent than complements without it; in written language the situation is the other way around. This leads Leech et al. (ibid., 191) to the conclusion that "the grammaticalization processes discussed and illustrated above seem largely to be confined to written English." Increase especially in the frequency of the construction *HELP* + bare infinitive without the object NP is "largely confined to writing and formal and elaborate speech . . ." (Leech et al. 2009, 193). If the movement towards the grammaticalization of *HELP* is, indeed, advancing mainly in written language, it would be a rather rare case of language change; normally we would expect spoken language to change first. According to Hoffmann et al. (2008, 6), "it is widely assumed in linguistics that when something changes in a language, that change generally starts in the spoken rather than the written variety."

McEnery and Xiao (2005, 173–176), however, seem to strongly disagree with the views that *HELP* is undergoing grammaticalization. They make the valid point that unlike quasi-modals *NEED* and *DARE*, which derive from lexical verbs, *HELP* cannot act as a both main verb and a modal auxiliary:

(9) *Helped Mrs Arthur Goldberg sponsor the art gallery? (McEnery and Xiao 2005, 174)
(10) Did Mrs Arthur Goldberg help sponsor the art gallery?

[15] Leech et al. (2009, 190) acknowledge, however, that forms with especially "inanimate-abstract object" NPs can have the grammaticalized meaning as well. Thus they treat their example (20), "[t]he right person may just happen to come along, or it may be necessary to take certain steps to *help this happen*" with *help* having the object NP *this*, as a case of "semantic bleaching".

Example (9) is hence ill-formed, but (10), which is modified from (9) and in which *HELP* has *do*-support, is a sound sentence. Furthermore, in their data[16] McEnery and Xiao also find cases of *HELP* in its causative 'contribute to' meaning selecting complements with *to*. They agree that there has been a shift in the meaning of *HELP*, but disagree with the view that grammaticalization accounts for that shift (McEnery and Xiao 2005, 173–176).

This disagreement offers an interesting starting point for further research. One possible explanation for the disagreement could originate from the fact that the term *grammaticalization* is interpreted differently by different authors; what is hence needed is clearly defined terminology. Consider the following definition of *grammaticalization*:

> [Grammaticalization] . . . refers primarily to the dynamic, unidirectional historical process whereby lexical items in the course of time acquire a new status as grammatical, morpho-syntactic forms, and in the process come to code relations that either were not coded before or were coded differently. (Traugott and König 1991, 189)

To use the Saussurian and Chomskyan terminology introduced above, the concepts and relations in the brain form the abstract *langue*/*competence*; the ways in which these concepts and relations are coded into the language are a part of the *parole*/*performance*. What happens during the process of grammaticalization, then, is that some words or constructions that already exist in the language, i.e. instances of *parole*, start referring to concepts that they did not refer to before; these concepts belong to the *langue* domain. According to Heine et al. (1991, 151) grammaticalization often occurs because of "unfulfilled communicative needs, or by the presence of cognitive contents for which there do not exist adequate linguistic designations." This need not always be the case, however, since grammaticalized structures may develop even if there are old and

[16] McEnery and Xiao's (2005) data comes from the following corpora: LOB, Brown, F-LOB, Frown, the spoken component of the BNC, and the Corpus of Professional Spoken American English (CPSA).

established ways of referring to some cognitive contents (ibid.). It is widely assumed that the process of grammaticalization results in a given word or structure starting to refer to a more abstract concept (Heine et al. 1991, 156). Traugott has defined this process as a three-level shift from

> . . . meanings grounded in more or less objectively identifiable extralinguistic situations to meanings grounded in text-making (connectives, anaphoric markers, etc.) to meanings grounded in the speaker's attitude to or belief about what is said. . . . (Traugott and König 1991, 189)

The more grammaticalized 'contribute to' sense of *HELP*, is, perhaps, on the level of it being a "text-making" (ibid.) constituent. According to Mair (1995, 265) in cases where *HELP* occurs in the 'contribute to' sense, it could often be replaced by prepositions, for example. Consider the following examples from the FLOB corpus, cited in Mair (ibid., 264–265):

> (11) A plan to help 60,000 jobless youngsters find work each year will be rejected by the TUC conference today.
> (12) This [additional money] is used to help pay the clergy and bills like heating and lighting.

According to Mair (ibid.), "[a] 'plan to help youngsters find jobs' is nothing more specific than a 'plan *for* them to find jobs'; and money used to 'help pay' for something is just money used *for* paying something." These paraphrases by prepositions suggest that the role that *HELP* has is, like the role of prepositions, to make connections between different parts of the sentence. This puts it on the level of a "text-making" (Traugott and König 1991, 189) constituent that has a connective role.

These changes in the reference of a word or construction explain the semantic side of the process of grammaticalization. Indeed, some authors have suggested a division between the semantics of grammaticalization and the impact that semantics has, then, on

the structure of a language. The semantic aspect of grammaticalization has been described in terms of semantic bleaching, depletion, weakening, or desemanticization (Heine et al. 1991, 155–156). The process has a certain input, referred to as the source concept or the source proposition, and an output, which, by definition, is more abstract than the input (ibid., 151–156). Source concepts are typically inherent to everyday human experiences; they are "'concrete' reference points for human orientation which evoke associations and are therefore exploited to understand 'less concrete' concepts" (Heine et al. 1991, 152).

Grammaticalization, then, has a cognitive base. Words start referring to more abstract concepts through an extension of meaning. Semantic changes may lead to the reanalysis of a given structure; a lexical word is reanalyzed as a grammatical word (Heine et al. 1991, 167–168). As Heine et al. note (ibid., 168), "[t]here is a remarkable amount of iconity between cognitive and linguistic patterning." This means that similar cognitive patterns are coded into the language using similar structures. Indeed, Mair (1995, 265) has argued that the "fossilising verbal form [of *HELP*] is gradually being reshaped as a semi-auxiliary or catenative verb, or even as a preposition or a conjunction." This, when taken together with the advancing bare infinitive constructions, could be treated as evidence suggesting that *HELP* is undergoing a process of grammaticalization in the syntactic sense of the term, too.

Conclusions based on earlier treatments of the grammaticalization of *HELP* are difficult to make. Perhaps the meaning of *HELP* "is undergoing semantic bleaching" (Leech et al. 2009, 190) and in some environments its "meaning is so general and abstract that it approaches those typically associated with grammatical categories" (Mair 2002,

124). However, in terms of syntax, the grammaticalization seems to be confined to the omission of the infinitive marker *to*, which typically occurs with modals in English. Unlike *NEED* and *DARE*, however, *HELP* cannot function as a modal auxiliary and needs *do*-support in interrogatives, for example, as was exemplified in (9) and (10) above. This would suggest that the syntactic status of *HELP* has not, as of today, gone through any exhaustive process of syntactic reanalysis.

4.4. Semantic distinction: the distance principle

Some authors have suggested that the choice between the *to*-infinitive and the bare infinitive is conditioned by a semantic distinction. Dixon (1991, 199; 2005, 201) has argued that the to-infinitive involves more direct causation than the bare infinitive. In (13) John assists Mary, but in (14) he takes part in the eating:

> (13) John helped Mary to eat the pudding (by guiding the spoon to her mouth, since she was still an invalid)
> (14) John helped Mary eat the pudding (he ate half) (Dixon 2005, 201)

The argument that there is a difference in meaning between the two infinitival complements is in line with Bolinger's Principle (Bolinger 1968, 127), according to which "a difference in syntactic form always spells a difference in meaning."

A careful examination of Dixon's analysis of the semantic distinction between the examples (13) and (14) above reveals, in fact, that the difference can be seen as a manifestation of a more general principle in linguistics, namely the distance principle (Rohdenburg 1996, 159). The principle has been defined in the literature as follows:

> The syntactic distance between two expressions corresponds to the semantic distance between them. (ibid.)

In (14) above the infinitive marker *to* has been omitted; thus there is a shorter syntactic distance between the subject of the higher clause, i.e. *John*, and the action described in the infinitive clause. According to the distance principle we can assume, then, that the semantic distance between the two is, correspondingly, shorter than in the variant with *to*. This gives rise to the interpretation that *John* may have eaten part of the pudding himself.

Other authors, however, do not place emphasis on this semantic distinction (Leech et al. 2009, 188) or reject the view that there is a clear semantic distinction between the patterns (McEnery and Xiao 2005, 169–173).

4.5. The *horror aequi* principle

The postulated *horror aequi* principle has been defined by Rohdenburg (2003, 236) as follows:

> . . . the *horror aequi* principle involves the widespread (and presumably universal) tendency to avoid the use of formally (near-) identical and (near-)adjacent . . . grammatical elements or structures. . . . (ibid.)

Relating this to the infinitival complements of *HELP*, the *horror aequi* principle potentially favours the bare infinitive when the matrix verb itself occurs in the *to*-infinitive. In other words, patterns of the type *to help* + *to*-infinitive might be avoided because of *horror aequi* effects. According to the principle, it is unlikely that *to* would be inserted before *meet* in example (15) below:

> (15) To help meet these objectives the Home Secretary appointed a Civil Emergencies Adviser. (Leech et al. 2009, 188)

Rohdenburg himself confirms this hypothesis by studying the complements of the verb forms *to help* and *helped*/*helps* in several corpora. According to his findings, *to*-infinitive clause complements were proportionally clearly less frequent after *to help* than *helped*/*helps* (Rohdenburg 2006, 157–158). This tendency is also noted by McEnery and Xiao (2005, 180–181). To counterpose this, Leech et al. (2009, 188) state that "instances of *to help to* are not at all rare, however. . . ." The effect that *horror aequi* has, then, is more "statistical . . . than categorical-structural" (ibid.). Incidentally, in patterns where *help* itself is preceded by *to*, an intervening (complex) NP could potentially cancel out the *horror aequi* effects found in *to help* + *to*-infinitive. The first and second *to*, which are fairly close to each other in the pattern *to help* + *to*-infinitive, are moved further away from each other by the intervening NP object in the pattern *to help* + NP + *to*-infinitive. Lind (1983, 266), for example, did not find any tokens of the type *to help* + *to*-infinitive in his data, but some of the type *to help* + NP + *to*-infinitive were attested. The impact of *horror aequi* on complement selection will be analyzed in this thesis with the help of corpus data in due course.

4.6. The complexity principle

The complex ways in which we process language in our brain is studied in the interdisciplinary field of psycholinguistics (or cognitive linguistics). Research in the area has shown that certain constructions are cognitively more complex to process than others. Rohdenburg (1996, 149–175) has developed this idea further and shown that the

cognitive complexity in the processing of certain constructions can have an impact on complement selection. This tendency has been expressed as the complexity principle:

> In the case of more or less explicit constructional options the more explicit one(s) will tend to be preferred in cognitively more complex environments. (Rohdenburg 2006, 147)

Several different types of constructions may constitute a cognitively complex environment. These include, for example, structural discontinuity (insertions) and passive transformations. What the complexity principle presupposes, of course, is that we must be able to decide on the hierarchy of explicitness between competing constructional options (Rohdenburg 1996, 152). As a rule of thumb, Rohdenburg (ibid.) suggests that "the more explicit variant is generally represented by the bulkier element or construction." For the present study this means that the *to*-infinitive can be regarded as the more explicit variant because it is the "bulkier" of the two constructions.

As regards the infinitival complements of *HELP*, there are several constructions that may constitute a cognitively complex environment. First, any insertions between *HELP* and the infinitival clause can be regarded as creating a complex environment (Rohdenburg 1996, 159). Consider examples (16) and (17) below, taken from Rohdenburg (ibid.), where an adverbial phrase *as well as I could* has been inserted:

(16) I helped (them) as well as I could to wash up.
(17) ? I helped (them) as well as I could wash up.

The insertion of an adverbial phrase clearly reduces the acceptability of the construction with the bare infinitive. Rohdenburg (ibid.) notes that the acceptability of the bare infinitive decreases as the complexity of the insertion increases. Incidentally, in environments such as (17) above, the distance principle that was presented in section 4.4 above is in contradiction with the complexity principle. The subject *I* is clearly directly

involved in the washing-up; hence, according to the distance principle, we might expect the bare infinitive to be used. However, the complexity principle seems to exert more power on the complement selection. As Rohdenburg (1996, 160) states, "in this complex environment the semantic distinction motivated by the distance principle seems to have been suspended."

In a similar vein to adverbial phrase insertions, it could also be hypothesized that the insertion of a complex NP object between *HELP* and the infinitive clause might constitute a complex environment in which the *to*-infinitive is more likely to occur than the bare infinitive.[17] Indeed, Rohdenburg (1996, 173) lists "(the presence of) more or less complex surface objects preceding finite and nonfinite clauses" as one of the factors that often triggers the choice of the more explicit complement. Elicitation tests on native speakers have shown that NP objects with premodification are more likely to choose the *to*-infinitive clause than NP objects that consist of a single pronoun (Quirk and Rusiecki 1982, 386–387). McEnery and Xiao (2005, 178–179) studied the impact of the number of the intervening words between the matrix verb *HELP* and infinitive clause.[18] Their findings indicate that number of the intervening words does not have a statistically significant influence on the choice between the two types of infinitive construction (ibid.).

The complexity principle may also exert power on complement selection in extraction environments:

(18) This is the system that he helped (to) establish [t]. (Rohdenburg 2006, 152)

[17] Furthermore, as was noted in Section 4.5 above, an intervening NP could potentially cancel out *horror aequi* effects in patters where *help* is preceded by *to*.

[18] McEnery and Xiao (2005, 178-179) do not make a difference whether the intervening words constitute an NP object or some other type of insertion.

In example (18) above the NP *the system* has been extracted from its object position after *establish* in the lower clause. The trace that it has left behind is marked with a [t]. The canonical version of (18) would be (19):

(19) He helped to establish the system.

Based on evidence from corpora, Rohdenburg (2006, 152) states that "[a]t every chronological stage, the erosion of the marked infinitive is clearly delayed in the more complex extraction contexts." Furthermore, negation can have its role in creating a complex environment. To support this, Leech et al. (2009, 187) note that "a negated infinitive seems to favour the use of *to*."

4.7. Different forms of *HELP*

With the help of large (and possibly annotated) corpora it is possible to compare the patterns of usage between different forms of a verb. As Sinclair (1991, 8) notes, this sort of study often reveals that different verb forms typically select different complementation patterns. Lind (1982, 265–269) studied the impact of different forms of *HELP* on the complement selection. According to his findings, *to* was likely to be omitted when it followed the verb form *help*. This was true irrespective of whether *help* was preceded by *to* or not. All other verb forms, *helping*, *helps*, and *helped*, favoured the *to*-infinitive. *Helping* seemed to be most resistant to the bare infinitive.

5. Corpus data analysis

After the theoretical part of this thesis I will now move on to classifying and analysing corpus data. As has been noted above, these data come from corpora representing written and published British English prose from the early 18th century to present-day English: the CLMET and the Imaginative prose section of the BNC. Chapter 5 consists of the analysis of each of the sub-corpora, working its way from the oldest data to the newest. In this chapter the approach will be synchronic to each sub-corpora; no diachronic comparisons are made at this point. In Chapter 6, in which the findings of this thesis will be presented, diachronic comparisons between different sets of data will be made.

As was noted above, the CLMET is not annotated for grammatical classes; hence each sub-section of the corpus was searched for each of the four forms of *HELP*, i.e. *help*, *helps*, *helped*, and *helping*. These searches collect all the instances of those word forms irrespective whether they are verbal, nominal, or adjectival. First, all non-verbal instances of *HELP* were manually removed from the data. Furthermore, since this study concentrates on the infinitival complements of *HELP*, those verbs which do not select an infinitival complement were discarded as being irrelevant for my purposes. The searches *help*, *helps*, *helped*, and *helping* collected altogether 327, 914, and 1666 tokens in CLMET parts 1, 2, and 3, respectively. After manual inspection of all of these tokens, it turned out that 30, 83, and 217 of these are verbs selecting one of the four possible infinitival complement patterns. Based on these numbers the precision of the queries can be calculated: the precision was 9.2%, 9.1%, and 13.3% for the CLMET parts 1, 2, and 3, respectively. When the CLMET is considered as a whole, the queries collected altogether

2907 tokens from all three sub-sections of the corpus, of which 330 are relevant for this study. Hence the joint precision of all the searches in the CLMET was 11.4%.

As concerns the Imaginative prose section of the BNC, a more elaborate set of queries was employed to extract data with comfortable precision and recall from the corpus. These search strings were described above in Section 2.5.2. After the manual inspection of the data from the BNC, a comfortable number of 1265 relevant tokens survived for further analysis. Hence altogether 1595 authentic instances of *HELP* followed by an infinitival complement were collected for this study. The present chapter and the one following it concentrate on analyzing these relevant tokens.

In each subsection of the present chapter I will first simply chart the distribution of the four infinitival complementation patterns. Secondly, the correlation between different patterns and the simplified senses established in Section 3.1 will be charted and authentic language samples given. Lastly, in each subsection I will make some remarks on *horror aequi* constraints, insertions, extractions, and passive constructions, all of which can have an impact on complement selection.

5.1. CLMET 1: 1710–1780

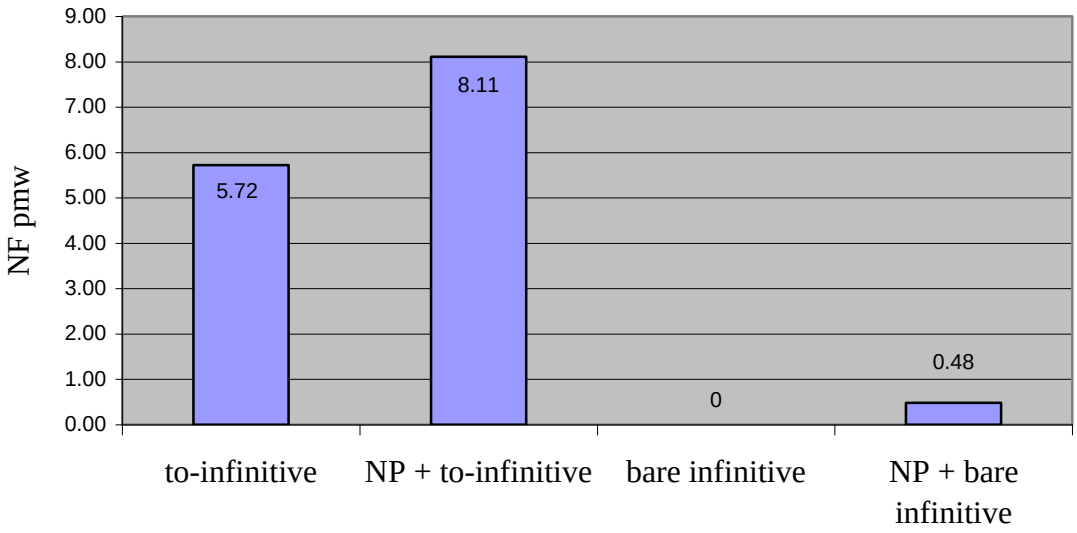
30 instances of the verb *HELP* selecting one of the infinitival complements were found in the CLMET 1. As the sub-corpus consists of 2,096,405 words (De Smet 2005, 73), the normalized frequency of all the instances of *HELP* selecting an infinitival complement in the sub-corpus is 14.31 pmw. Table 8 below summarises the distribution of the infinitival complements across the four verb forms:

Table 8. Infinitival complements of *HELP* in the CLMET 1. Size 2,096,405 words.

Complement	*help*	*helps*	*helped*	*helping*	**Total**	% of all comps	NF pmw
to-infinitive	8	1	2	1	**12**	40.0 %	5.72
NP + *to*-infinitive	10	3	3	1	**17**	56.7 %	8.11
bare infinitive	-	-	-	-	**0**	0 %	0
NP + bare infinitive	1	-	-	-	**1**	3.3 %	0.48
Total	19	4	5	2	**30**	100 %	14.31

As can be seen in Table 8, complements with the *to*-infinitive dominate in the data from the CLMET 1. A mere one token of the type *help* + NP + bare infinitive was attested. No instances of the bare infinitive without the object NP were collected. Figure 3 visually presents the normalized frequencies of the four infinitival complement patterns:

Figure 3. CLMET 1. The normalized frequencies of complements pmw.



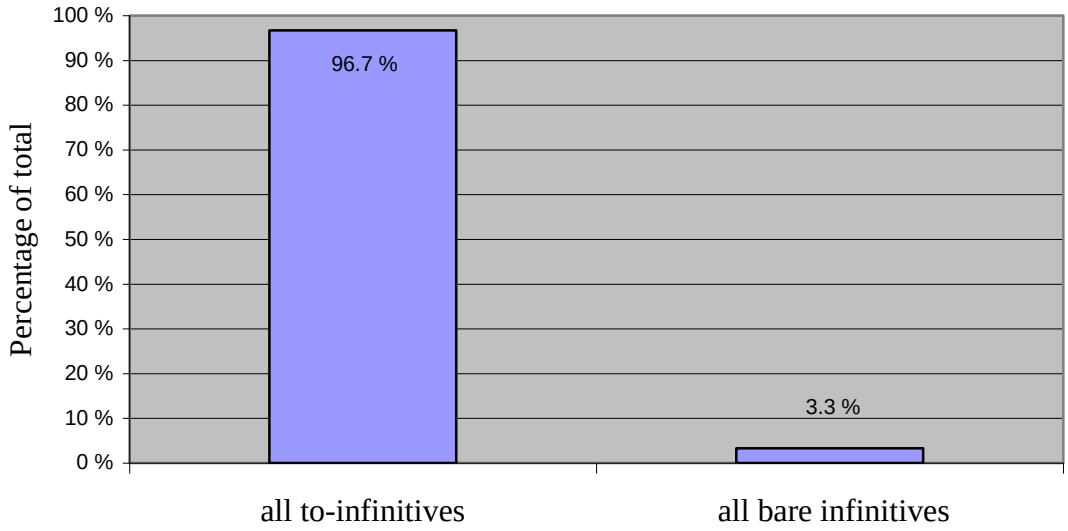
Since one of the aims of this study is to compare changes in the distribution of the two main types of complements, the *to*-infinitives and bare infinitives, it is useful to collapse the four infinitival complement patterns into two main types. The *to*-infinitive and NP + *to*-infinitive patterns are collapsed into one main pattern representing all the *to*-infinitive

complements; the bare infinitive and NP + bare infinitive patterns are put under one main pattern representing all the bare infinitive complements. This allows fluent comparisons to be made in the distribution of the two main types between different sets of data.

Table 9. CLMET 1. Distribution of the two main types as a percentage of all.

All *to*-infinitives	All bare infinitives	All *to*-infinitives as % of total	All bare infinitives as % of total
29	1	96.7 %	3.3 %

Figure 4. CLMET 1. Distribution of all to-infinitives and all bare infinitives.



CLMET 1 contains almost exclusively infinitival complements selecting either of the constructions with the *to*-infinitive. Bare infinitives are extremely rare with mere one token in the data.

Section 3.1 discussed the correlation between sense and structure. The *OED* analysis suggested that the complements without the intervening object NP could manifest either the 'aid' sense, which would derive from the omission of the object NP in the prototypical 'A helps B to do C', or be instances of the more desemanticized 'contribute to' sense. Patterns with the object NP, as in 'A helps B to do C', on the other hand, are, according to the *OED*, more likely to code the prototypical 'aid' sense.

Table 10 charts the correlation between different patterns and their compatibility with the 'aid' and 'contribute to' senses in the CLMET 1:

Table 10. CLMET 1. Complement/sense mapping.

Complement	**Sense** 'aid'	**Sense** 'contribute to'
to-infinitive	3	9
NP + *to*-infinitive	17	-
bare infinitive	-	-
NP + bare infinitive	1	-
Total	**21**	**9**

There are 12 instances of *HELP* followed by a *to*-infinitive complement in the CLMET 1. Each of the four verb forms were found followed by the *to*-infinitive. A careful analysis of the corpus data suggests that the pattern *HELP* + *to*-infinitive is more likely to code the 'contribute to' sense. Although the *OED* acknowledged the 'contribute to' sense with the *HELP* + *to*-infinitive pattern, this finding was not expected on the basis of the review of earlier literature; Mair (2002, 123–126), for example, explained the overall increase in the frequency of *HELP* with an infinitival complement clause in the 20th century in terms of grammaticalization of its meaning from the prototypical 'aid' sense to the more grammaticalized 'contribute to' sense. The fact that the 'contribute to' sense is found in the data from the 18th century and especially with the *HELP* + *to*-infinitive pattern is somewhat surprising. Although in some cases the division into two clear-cut semantic classes is difficult to make, I singled out three instances of the 'aid' sense and nine of the 'contribute to' sense with the pattern *HELP* + *to*-infinitive. Examples (20) to (24)

illustrate patterns with the *to*-infinitive, with two examples chosen to exemplify both the 'aid' and 'contribute to' senses:[19]

(20) She expressed no sort of concern for her brother, who ran about in the cold, without his periwig, and worked with the most astonishing agility, in helping to disentangle the horses from the carriage . . . (Smollet 1751, *The Expedition of Humphrey Clinker*)
(21) A better commander I'd never desire to serve; and who knows but I may help to set up thy standing rigging in another world? (Smollet 1751, *The Adventures of Peregrine Pickle*)
(22) I cannot avoid mentioning here a circumstance in portrait painting which may help to confirm what has been said. (Reynolds 1769–1776, *Seven Discourses on Art*)
(23) Dryden alone (what wonder?) came not nigh / Dryden alone escaped this judging eye / But still the great have kindness in reserve / He helped to bury whom he helped to starve. (Pope 1733–1734, *An Essay on Man*)

In (20), where *helping* has a quite nouny taste to it, the subject *brother* is clearly aiding someone. In (21) a personal pronoun *you* could easily be inserted after *help*, as in *I may help you to set up thy standing*. . . . These two examples are quite clear cases of the 'aid' sense. In (22) *help* could, in fact, be quite easily replaced with *contribute to*: . . . *which may contribute to confirming*. . . . (23) is more difficult to analyze because of the poetic tone of the example. It was analyzed, however, as a case of the 'contribute to' sense because it is unlikely that the subject is concretely aiding someone or something, but rather playing some abstract role in the process.

As can was expected on the basis of the *OED* analysis, all of the instances of the pattern *HELP* + NP + *to*-infinitive were found to manifest the prototypical 'aid' sense. Again, the pattern was found occurring after each of the four inflected forms of *HELP*. Examples (24) and (25) exemplify this pattern:

[19] References to authentic samples from the CLMET are given as the name of the author, year of publication, and title of the work

(24) I agree to your removing to an 'hotel garni'; the Abbe will help you to find one, as I desire him by the inclosed, which you will give him. (Chesterfield 1746–1771, *Letters to his Son*)

(25) . . . but this, indeed, is the great optic-glass helping us to discern plainly almost all that passes in the minds of others . . . (Fielding 1751, *Amelia*)

In both (24) and (25) there is a clear sense of 'A helps B to do C'. The data from the CLMET 1 suggest that the object NP clearly triggers the 'aid' interpretation.

As was noted above, there were no instances of the bare infinitive without the object NP in the CLMET 1. The single instance of the pattern *HELP* + NP + bare infinitive was found after the verb form *help*:

(26) I should be glad you was present at the nuptials, to help me throw the stocking, and perform other ceremonies peculiar to that occasion. (Smollett 1771, *The Expedition of Humphrey Clinker*)

This token manifests the 'aid' sense, as was predicted by the presence of the object NP.

There were two instances where *help* was preceded by the infinitive marker *to* in the CLMET 1. Interestingly, as can be seen in (26), the only instance of the NP + bare infinitive complement in the data was found after one of these. This can be treated as evidence that the author was, perhaps unconsciously, avoiding "the use of formally (near-) identical and (near-)adjacent . . . grammatical elements or structures . . ." (Rohdenburg 2003, 236). However, the other instance of *to help* was followed by an NP + *to*-infinitive complement:

(27) I had once lost my portmanteau from behind my chaise, and twice got out in the rain, and one of the times up to the knees in dirt, to help the postilion to tie it on, without being able to find out what was wanting. (Sterne 1768, *A Sentimental Journey through France and Italy*)

This instance of *to help* + NP + *to*-infinitive is not, however, a very severe violation of the *horror aequi* principle. As was discussed in Section 4.5, an intervening object NP between *to help* and the *to*-infinitive in the complement, which in this case is two words

long, moves the two infinitive markers further away from each other, potentially cancelling out *horror aequi* effects. There were no sequences of the type *to help* + *to*-infinitive in the data.

Some instances of potentially cognitively complex environments were found in the data. Two insertions were attested: *in the least* in (28) and *but* in (29):

(28) . . . but my perfect conviction of this does not help me in the least to discern . . . (Burke 1775, *On Conciliation with America*)
(29) . . . help me but to find a trap-door that must be hereabout . . . (Walpole 1764, *The Castle of Otranto*)

Both were attested with the *to*-infinitive. In one instance the object NP had been extracted:

(30) And me, the Muses helped to undergo it . . . (Pope 1733–1734, *An Essay on Man*)

This token was analyzed as an NP + *to*-infinitive complement, because the canonical version of the sentence would read as *the Muses helped me to undergo it*. No passive constructions were found in the data from the CLMET 1.

5.2. CLMET 2: 1780–1850

The searches *help*, *helps*, *helped*, and *helping* collected altogether 914 tokens in the CLMET 2. All these were inspected manually to weed out all undesired tokens. After this procedure, 83 instances of *HELP* selecting one the infinitival complementation patterns remained. The CLMET 2 consists of 3,739,657 words (De Smet 2005, 75), which gives the normalized frequency of 22.19 pmw for all the instances of *HELP* selecting an

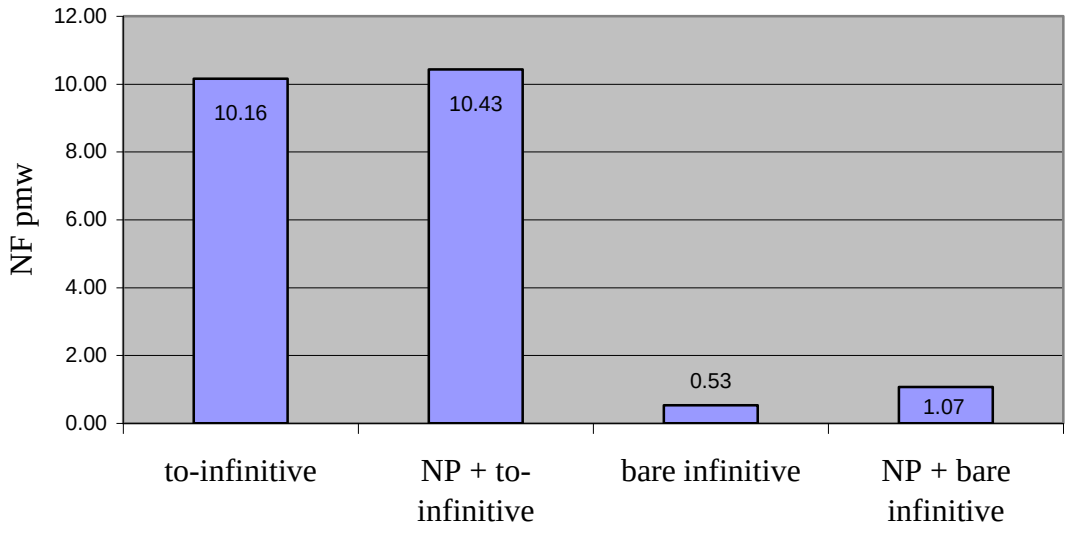
infinitival clause complement in the sub-corpus. Table 11 below summarises their distribution:

Table 11. Infinitival complements of *HELP* in the CLMET 2. Size 3,739,657 words.

Complement	*help*	*helps*	*helped*	*helping*	**Total**	% of all comps	NF pmw
to-infinitive	16	2	17	3	**38**	45.8 %	10.16
NP + *to*-infinitive	29	-	8	2	**39**	47.0 %	10.43
bare infinitive	2	-	-	-	**2**	2.4 %	0.53
NP + bare infinitive	4	-	-	-	**4**	4.8 %	1.07
Total	51	2	25	5	**83**	100 %	22.19

To-infinitive and NP + *to*-infinitive complements dominate in the data from the years 1780 to 1850. However, some tokens with the bare infinitive were collected. Two of these were of the type *HELP* + bare infinitive and four of the type *HELP* + NP + bare infinitive. Figure 5 presents the normalized frequencies of the patterns as diagrams:

Figure 5. CLMET 2. The normalized frequencies of complements pmw.



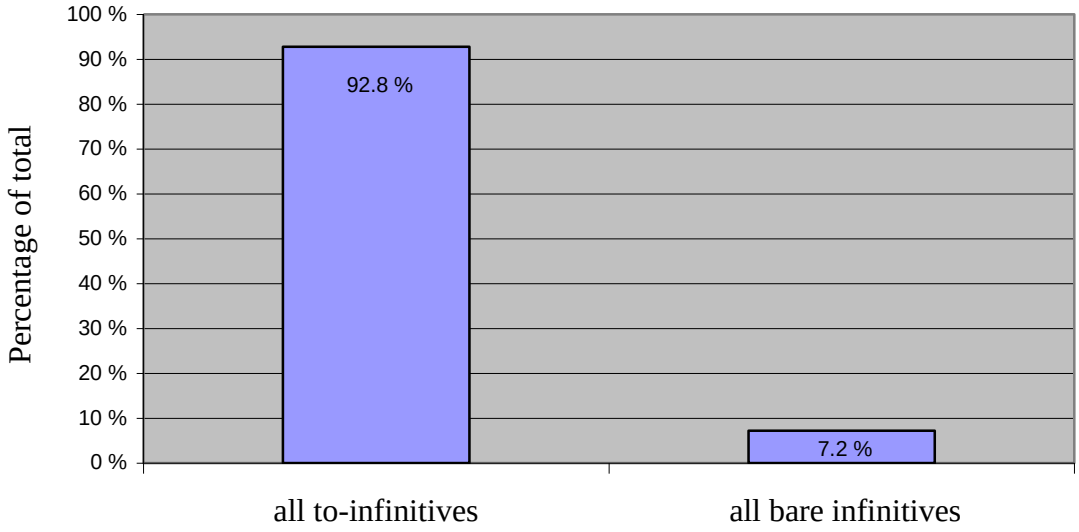
As in the previous section, the four infinitival complement types are collapsed into two main types for ease of comparison: in Table 12 and Figure 6 below the *to*-infinitive and

NP + *to*-infinitive complements are put under the label of *all to-infinitives*; the umbrella term for the bare infinitive and NP + bare infinitive complements is *all bare infinitives*.

Table 12. CLMET 2. Distribution of the two main types as a percentage of all.

All *to*-infinitives	All bare infinitives	All *to*-infinitives as % of total	All bare infinitives as % of total
77	6	92.8 %	7.2 %

Figure 6. CLMET 2. Distribution of all *to*-infinitives and all bare infinitives.



Infinitival complements with the marker *to* are far more common than variants without it in the data from the CLMET 2. Constructions with the bare infinitive remain rare in the data with six tokens as opposed to 77 tokens of complements with the infinitival marker *to*.

Although the division into two clear-cut semantic classes is somewhat problematic and always involves some subjective decisions based on the analyst's intuition, this was, again, attempted. Table 13 maps the correlation between different patterns and senses in the CLMET 2:

Table 13. CLMET 2. Complement/sense mapping.

Complement	**Sense** 'aid'	**Sense** 'contribute to'
to-infinitive	14	24
NP + *to*-infinitive	39	-
bare infinitive	2	-
NP + bare infinitive	4	-
Total	**59**	**24**

To-infinitive complements were found after each of the four inflected forms of *HELP*. Of the 38 *to*-infinitive complements in the CLMET 2 about two thirds manifest the 'contribute to' sense, leaving one third for the 'aid' sense. As was noted in the previous subsection, this sense was not expected to be prominent in the historical data. Examples (31) to (34) illustrate the *HELP* + *to*-infinitive pattern, first two being examples of the 'aid' sense and the latter two of the 'contribute to' sense:

(31) . . . there is no one but me to assist her. I can help to get the articles together, but I do not understand the art of stowing them into the boxes . . . (Brontë, Anne 1848, *The Tenant of Wildfell Hall*)
(32) . . . and Dolly, hanging about him in all kinds of graceful winning ways, helped to button and buckle and brush him up and get him into one of the tightest coats that . . . (Dickens 1841, *Barnaby Rudge*)
(33) . . . the Maragatos, like true men of the north, delight in swilling liquors and battening upon gross and luscious meats, which help to swell out their tall and goodly figures. (Borrow 1842, *The Bible in Spain*)
(34) . . . the welcome breeze helps to refresh me, and ease the tightness at my breast! (Hazlitt 1821–1822, *Table Talk*)

In (31) an object pronoun *her* could easily be inserted after *help* and in (32) the subject *Dolly* is clearly helping someone. These are fairly clear cases of the 'aid' sense. In examples (33) and (34), however *HELP* has the sense of furthering or enabling and hence they were classified as instances of the 'contribute to' sense.

Again, all instances of the *HELP* + NP + *to*-infinitive pattern were found to manifest the prototypical ‘aid’ sense. These were found after the verbal forms *help*, *helped*, and *helping*. Leech et al. (2009, 190) noted that patterns with especially “inanimate-abstract object” NPs could be instances of the grammaticalized sense of the verb, too, but no such instances were found in the data from the CLMET 2. Examples (35) and (36) illustrate this pattern.

(35) Nelly, help me to convince her of her madness. (Brontë, Emily 1847, *Wuthering Heights*)
(36) When Edmund came, he helped them to move the great stone from off the poor animal . . . (Edgeworth 1796–1801, *The Parent’s Assistant*)

In both examples there is a sense of ‘A helps B to do C’.

In the CLMET 2 the *HELP* + bare infinitive pattern emerged with two tokens, both following the verb form *help*:

(37) After they left us, I went with my Mother to help look at some houses in New King Street (Austen 1796–1817, *Letters to her Sister Cassandra and Others*)
(38) No, Missy; I tell Massa Tommy to help carry cocoa-nut leaves, and then he go away directly. (Marryat 1841, *Masterman Ready*)

The review of earlier literature on *HELP* especially in Section 4.3 suggested that the increase especially in the frequency of the pattern *HELP* + bare infinitive is promoting the interpretation that the verb is being grammaticalized. Perhaps unexpectedly, both (37) and (38) turned out to be instances of the ‘aid’ sense. In (37) the speaker is helping her mother and in (38) there is a clear sense of ‘A helps B to do C’, too.

All four tokens of the type *HELP* + NP + bare infinitive followed the form *help* and were cases of the ‘aid’ sense:

(39) Here! just help me lift the little sofa near the fire, will you, Uncle Sol . . . (Dickens 1848, *Dombey and Son*)

(40) What's up that he can't come home and help me nurse? (Gaskell 1848, *Mary Barton*)

Again, the object NP seems to trigger the 'aid' interpretation.

As can be seen in (37) and (38) above, both instances of the newly emerged pattern *HELP* + bare infinitive were found in constructions where *help* was preceded by *to*. Although two tokens are not enough to draw any far-reaching conclusions, the *horror aequi* principle might be an influencing factor. None of the four tokens with the NP + bare infinitive complement were, however, preceded by *to help*; two of these were exemplified in (39) and (40) above. The omission of the infinitive marker could be accounted for by the distance principle: in both (39) and (40) above there is a strong sense that the subject is directly helping the person referred to by the object NP. The shorter syntactic distance between the matrix verb and the action described in the complement clause could hence be accounted for by the shorter semantic distance between them. As regards violations of the *horror aequi* principle, there were two instances of *to help* + *to*-infinitive construction in the CLMET 2:

(41) I added that I had brought a small stock of Bibles and Testaments to Evora, which I wished to leave for sale in the hands of some respectable merchant, and that if he were anxious to help to lay the axe to the root of superstition and tyranny, he could not do so more effectually than by undertaking the charge of these books. (Borrow 1842, *The Bible in Spain*)

(42) . . . and left all us, who had been taken prisoners by the French, in the vessel, to help to work her into port, as the captain did not wish to part with any more men of his own. (Marryat 1841, *Masterman Ready*)

Instances of *to help* + NP + *to*-infinitive, where the two infinitive markers are distanced by the intervening object NP, were more frequent with nine instances. Two of these are presented in (43) and (44):

(43) After he was gone, Hannah Rogers, one o' th' neighbours came in and wanted me to help her to wash. (Brontë, Anne 1847, *Agnes Grey*)
(44) . . . and the duty of the happy is to help the suffering to bear their woe. (Gaskell 1848, *Mary Barton*)

As has been noted above, these are not, perhaps, as severe violations as instances of *to help* + *to*-infinitive.

Three insertions were found in the data. All of these favoured infinitives with *to*:

(45) . . . and I thought I would never leave a thing about, as might help, ever so little, to convict him. (Gaskell 1848, *Mary Barton*)
(46) . . . and she should help me, first to find out a picture dealer in some distant town; then, through her means . . . (Brontë, Anne 1848, *The Tenant of Wildfell Hall*)
(47) Then you won't help me, Job, to prove him innocent? (Gaskell 1848, *Mary Barton*)

The data from the CLMET 2 backs up the statement that an insertion constitutes a cognitively complex environment in which the more explicit *to*-infinitive is favoured. One instance of a passive construction was attested and it occurred with the *to*-infinitive:

(48) His minute portions of Latin grammar, &c., were to be repeated over to him, till he chose to say he knew them, and then he was to be helped to say them.

Because the focus of this thesis is on argument structure and complementation, the passive construction *he was to be helped to say them* was related to its active version *somebody was to help him to say them* and hence treated as an instance of the pattern *help* + NP + *to*-infinitive.

5.3. CLMET 3: 1850–1920

The searches *help*, *helps*, *helped*, and *helping* collected altogether 1666 tokens in the CLMET 3. Again, these were analyzed manually. This left 217 instances of *HELP*

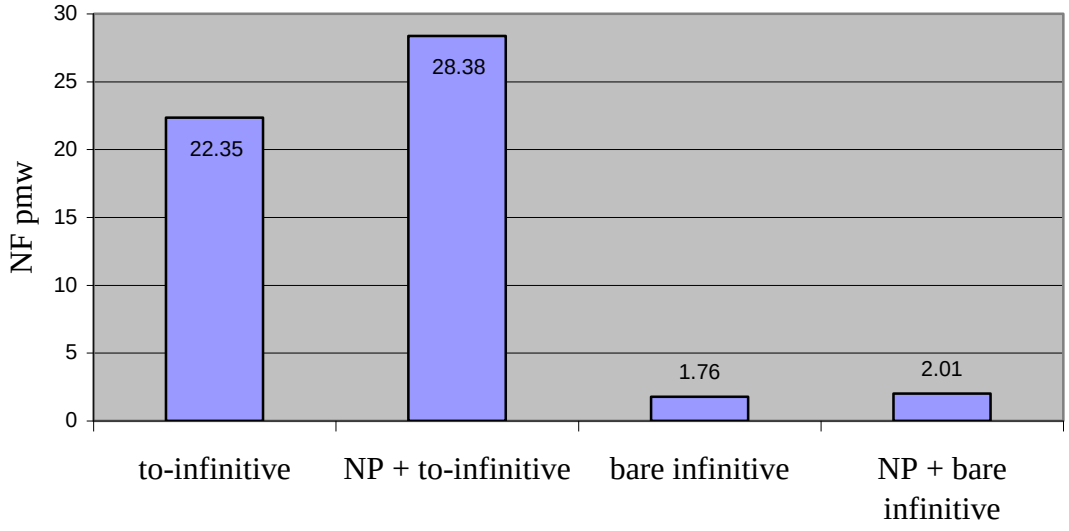
selecting an infinitival complement pattern for further analysis. The CLMET 3 is 3,982,264 words in size (De Smet 2005, 78), which gives the normalized frequency of 54.49 pmw for all the infinitival complementation patterns in the sub-corpus.

Table 14. Infinitival complements of *HELP* in the CLMET 3. Size 3,982,264 words.

Complement	*help*	*helps*	*helped*	*helping*	**Total**	% of all comps	NF pmw
to-infinitive	35	7	39	8	**89**	41.0 %	22.35
NP + *to*-infinitive	70	1	32	10	**113**	52.1 %	28.38
bare infinitive	5	-	2	-	**7**	3.2 %	1.76
NP + bare infinitive	8	-	-	-	**8**	3.7 %	2.01
Total	118	8	73	18	**217**	100 %	54.49

The numbers in Table 14 indicate that the *to*-infinitive and NP + *to*-infinitive complements were clearly favoured over the bare infinitive and NP + bare infinitive complements in the data from the CLMET 3. The normalized frequencies of each pattern are presented in Figure 7:

Figure 7. CLMET 3. The normalized frequencies of complements pmw.

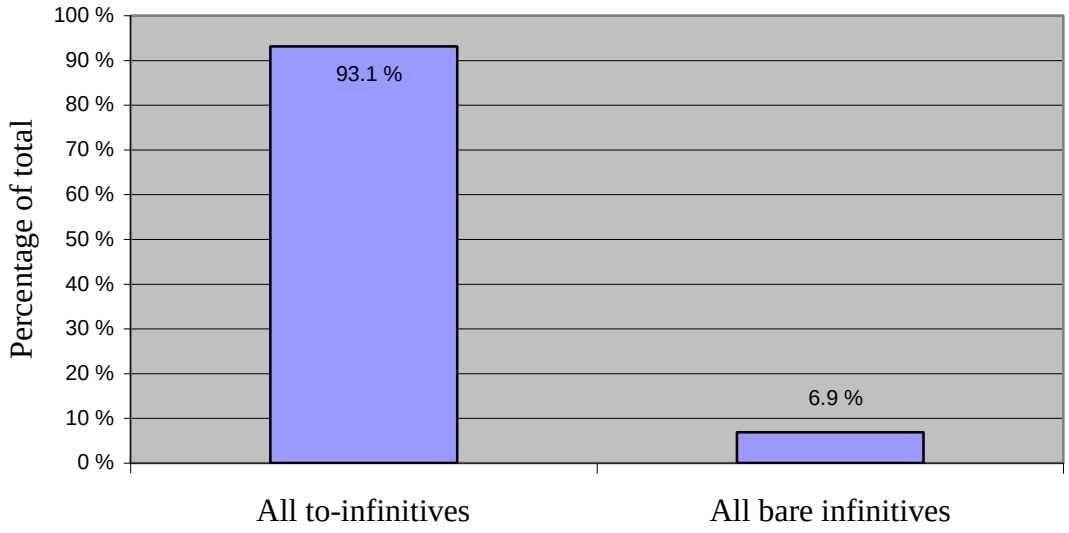


Again, *to*-infinitive and NP + *to*-infinitive complements are collapsed in one main type representing all *to*-infinitives; bare infinitive and NP + bare infinitive complements are treated under one main type of all bare infinitives:

Table 15. CLMET 3. Distribution of the two main types as a percentage of all.

All *to*-infinitives	**All bare infinitives**	**All *to*-infinitives as % of total**	**All bare infinitives as % of total**
202	15	93.1 %	6.9 %

Figure 8. CLMET 3. Distribution of all *to*-infinitives and all bare infinitives.



The data from the years 1850 to 1920 suggests that infinitival complements with the infinitive marker *to* were favoured over variants without it during the period under investigation. An infinitive clause with *to* was used more than nine times out ten instances of an infinitive clause controlled by *HELP*.

All 217 relevant tokens in the CLMET 3 were analyzed for semantic differences. Although an array of subtle semantic differences is bound to be present in a large set of data, all tokens were divided into the two simplified sense categories ‘aid’ and ‘contribute to’. The results of this analysis are presented in Table 16:

Table 16. CLMET 3. Complement/sense mapping.

Complement	**Sense** 'aid'	**Sense** 'contribute to'
to-infinitive	37	52
NP + *to*-infinitive	112	1
bare infinitive	7	-
NP + bare infinitive	8	-
Total	**164**	**53**

To-infinitive complements were found after each of the inflected forms of *HELP*. With 52 against 37 tokens, the majority of these were found to be instances of the 'contribute to' sense; again, the 'contribute to' sense was found already in the historical data and with the *HELP* + *to*-infinitive pattern. Examples (49)–(50) illustrate the 'aid' sense and (51)–(52) the 'contribute to' sense:

(49) Children! Help to look for him! Quick! (Carroll 1889, *Sylvie and Bruno*)
(50) Whatever William did Lanfranc no doubt at least helped to plan. (Freeman 1888, *William the Conqueror*)
(51) '. . . I should say that even the business serves a good purpose.' 'What purpose?' 'It helps to spread civilisation.' (Gissing 1891, *New Grub Street*)
(52) Some of our Auxiliaries have materially helped us in this way by distributing our literature at the seaside and elsewhere . . . thus helping to dispel the prejudice under which many persons unacquainted with the Army are found to labour. (Booth 1890, *In Darkest England and the Way out*)

In (49) and (50) there is a clear sense of aiding; in (50) the pronoun *him* could, in fact, quite easily be inserted after *helped* with little change to the meaning of the sentence. (51) and (52) are cases of the more desemanticized 'contribute to' sense of *HELP*. This sense can be characterized by the notions of furthering or promoting: in (51), the spread of civilization is promoted; in (52), the subject is contributing to dispelling some prejudices.

NP + *to*-infinitive complements were found after each of the inflected forms of *HELP*. With the exception of one token all of these were found to be instances of the 'aid' sense. Two tokens with the 'aid' sense are presented in (53)–(54) and the single instance of the 'contribute to' sense in (55):

(53) They [discussions between native North Americans], in some material respects, improve the sheep [native North Americans] within the fold; but they do not help them or incline them to leap out of the fold. (Bagehott 1869, *Physics and Politics*)[20]
(54) One of my cousins, who was a medical student, showed me a pistol, and helped me to fire it. (Gosse 1907, *Father and Son*)
(55) When he had reached Camden Town railway-station he was attracted by a coffee-stall; a draught of the steaming liquid, no matter its quality, would help his blood to circulate. (Gissing 1891, *New Grub Street*)

Though in (53) the language that is used is quite abstract, there is still a sense of aiding someone. In (54), a person is aided to fire a gun. These two are instances of the 'aid' sense. Interestingly, the pattern *HELP* + NP + *to*-infinitive was found to be closer to the 'contribute to' sense on one occasion, as exemplified in (55): a cup of hot coffee would promote the speaker's blood circulation. This interpretation might stem from the qualities of the object NP. The NP *his blood* is a non-human, non-reasoning entity, which cannot easily be aided to do something; the 'aid' interpretation would presuppose that the object NP is capable of wilfully striving to a goal, and hence can be aided to reach it. The NP *his blood* does not have these properties. In the vast majority of tokens of the type *HELP* + NP + *to*-infinitive, however, the object NP denotes a human entity that has a will of its own, which, perhaps, gives rise to the 'aid' interpretation.

[20] Though in (53) the complement of *help* might be a bit unclear because of another verb, *incline* is present, it is regarded as having an NP + *to*-infinitive complement. The infinitive clause *to leap out of the fold* belongs to the argument structures of both *help* and *incline* and complements their meanings. Instances similar to (53) in the data are considered relevant throughout this thesis.

Seven tokens of the type *HELP* + bare infinitive were found in the CLMET 3, five of which occurred after the verb form *help* and two after the verb form *helped*. Again, all these were found to manifest the ‘aid’ sense, which was not expected on the basis of the review of earlier literature:

> (56) . . . he fainted away, and his son, William Pitt, was called out of the House of Commons to help carry him away to his coach. (Younge 1873, *Young Folk’s History of England*)
> (57) “I helped bait up trawl ashore 'fore I could well walk,” he said. (Kipling 1897, *Captain Courageous*)

In both cases there is a clear sense of aiding someone. (57) is reported colloquial speech; although the statement was put under suspicion in the theory part of this thesis, the *OED* stated that the bare infinitive is nowadays “a common *colloq.* form” (*OED* s.v. *help* v. sense 5.a).

All eight tokens with an NP + bare infinitive complement were found after the verb form *help*. All of these manifested the ‘aid’ sense of the verb:

> (58) Amy had to help her mistress make herself as comely as she could be made without her best dress . . . (Bennett 1908, *The Old Wives' Tale*)
> (59) On Thursday I went round early in the morning to help Janie pack the baskets for the picnic. (Jerome 1909, *They and I*)

The object NPs are human entities normally referred to by a personal pronoun or a proper name. These qualities of the object NP might contribute to assigning them the ‘aid’ interpretation.

Four of the seven instances of the bare infinitive complements were cases where *help* occurred in the *to*-infinitive. The authors were, perhaps, avoiding the use of two infinitive markers in near-adjacent positions. One of the three instances where this was not the case was illustrated in (57), and the single instance of the bare infinitive

complement that occurred after the word form *help* that was not preceded by *to* is exemplified in (60):

> (60) "I say!" said Ogilvy; "help keep these idiots back. We don't know what's in the confounded thing, you know!" (Wells 1897, *The War of the Worlds*)

No obvious syntactic reason can be said to account for the omission of *to* in either (57) or (60). A semantic reason suggested in the literature that could account for the bare infinitive is the distance principle. Six of the eight instances of the NP + bare infinitive complements occurred in environments where *help* was preceded by *to*. Two of these are exemplified in (58) and (59) above. The two instances where *help* was not preceded by *to* are illustrated in (61)–(62):

> (61) Here, lend a hand, one of you, and help me pull out this young howling brute. (Hughes 1857, *Tom Brown's School Days*)
> (62) Let us help you put the things away. (Hughes 1857, *Tom Brown's School Days*)

Again, the distance principle might be a contributing factor to the omission of the infinitive marker *to*. In both (61) and (62) the subject is actively taking part in the action denoted in the infinitive clause. There were some violations of the *horror aequi* principle in the data:

> (63) Of course all the reapers came at night to the harvest-supper, and Parson Bowden to say the grace as well as to help to carve for us. (Blackmore 1869, *Lorna Doone*)
> (64) I learnt the latter language in order to help to teach it to him, but after five years of it he knew it as well as I did . . . (Haggard 1887, *She*)

There were five instances of the sequence *to help* + *to*-infinitive in the CLMET 3. Instances of the type *to help* + NP + *to*-infinitive are more common, but as has been noted, these might not be very severe violations of the *horror aequi* principle.

Three insertions were found in the CLMET 3. Again, all of these were attested with an infinitive with the marker *to*:

(65) . . . it must be confessed that some of these portraits scarcely help at all to explain the power of the players to whom they belonged. (Pater 1886–1890, *Essays from the Guardian*)
(66) But a manly resolve to recognize boldly that he had no longer a lover's interest in her, helped him occasionally to conceal a feeling. (Hardy 1874, *Far from the Madding Crowd*)
(67) O God, and the Saints, help me not to lose my way again! (Yonge 1870, *The Caged Lion*)

In (65) *at all* is inserted between the matrix verb and the infinitive clause; in (66) *occasionally* occurs between the object NP and the infinitive clause. (67) supports the statement made by Leech et al. (2009, 187), according to which "a negated infinitive seems to favour the use of *to*."

One extraction caught the analyst's eye. This was already recorded in (50), reprinted here as (68):

(68) Whatever William did Lanfranc no doubt at least helped to plan [t]. (Freeman 1888, *William the Conqueror*)

The object of *plan*, which belongs to the infinitival complement clause of *helped*, has been extracted to the sentence-initial position. The trace that it has left behind is marked with a [t] in (68). The canonical version of (68) would read as *Lanfranc no doubt at least helped to plan whatever William did*.

Six passive constructions were attested in the CLMET 3, three of which are recorded in (69)–(71):

(69) This child should be helped to live. (Meredith 1895, *The Amazing Marriage*)
(70) They were helped by these guides to recognize in wild Oriental visions direct statements regarding Napoleon III . . . (Gosse 1907, *Father and Son*)

(71) She was very much pleased at being helped to do what she felt to be right and kind . . . (Yonge 1865, *The Clever Woman of the Family*)

As was motivated in section 3.2, passive constructions are related to their active versions in the study of argument structure and complementation in this thesis. The active versions of (69)–(71) are presented in (72)–(74):

(72) Somebody should help this child to live.
(73) These guides helped them to recognize . . .
(74) She was very much pleased at somebody helping her to do . . .

On the basis of these active versions, tokens (69)–(71) were treated as instances of *help* + NP + *to*-infinitive, *helped* + NP + *to*-infinitive, and *helping* + NP + *to*-infinitive, respectively, and were included in these categories in Table 14 above. All six passive constructions in the sub-corpus were found to occur with the infinitive with *to*.

5.4. The BNC: the 1990s

A more elaborate set of queries was performed to extract data with comfortable precision and recall from the Imaginative prose section of the BNC. The results of these queries were analyzed manually to weed out any irrelevant tokens. This procedure resulted in 1265 relevant instances of *HELP* controlling one of the four infinitival clause complements. As the Imaginative prose section of the BNC is 16,496,408 words in size, the normalized frequency of all infinitival complements in the data is 76.68 pmw. Table 17 charts the distribution of the complements across the four inflected forms of *HELP*:

Table 17. Infinitival complements of *HELP* in the Imaginative prose section of the BNC. Size 16,496,408 words.

Complement	*help*	*helps*	*helped*	*helping*	**Total**	% of all comps	NF pmw
to-infinitive	75	17	101	44	**237**	18.7 %	14.37
NP + *to*-infinitive	178	19	126	51	**374**	29.6 %	22.67
bare infinitive	168	2	54	6	**230**	18.2 %	13.94
NP + bare infinitive	273	19	91	41	**424**	33.5 %	25.70
Total	694	57	372	142	**1265**	100 %	76.68

The numbers in Table 17 indicate that the complements without the object NP, the *to*-infinitive and bare infinitive complements, occur in the data with approximately equal frequency. The NP + bare infinitive complements are most frequent in the data followed closely by the NP + *to*-infinitive complements. The normalized frequencies of each pattern are presented in Figure 9:

Figure 9. BNC Imaginative prose. The normalized frequencies of complements pmw.

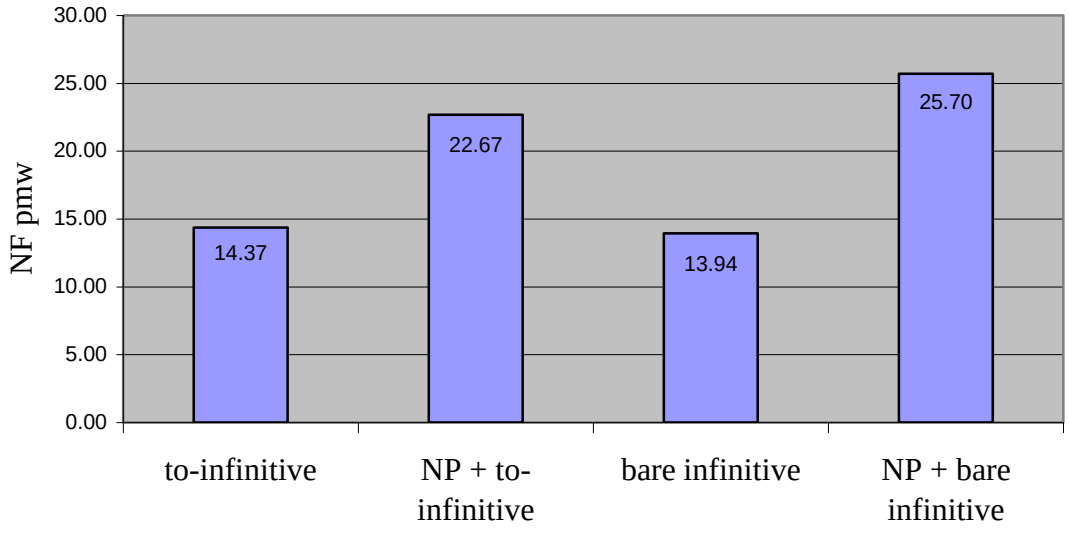
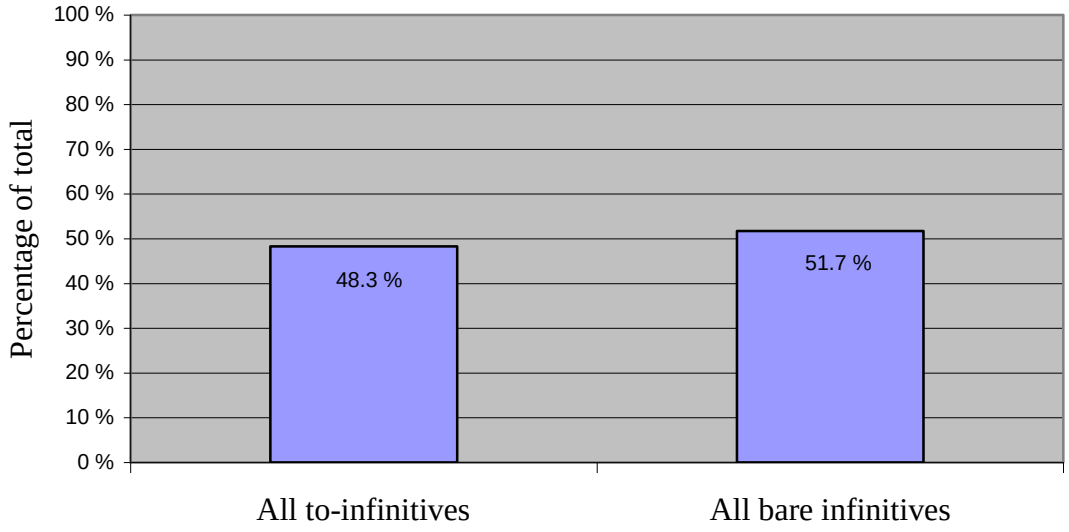


Table 18 and Figure 10 below chart the situation when the two main types, the infinitival complements with the marker *to* and the infinitival complements without it are compared:

Table 18. BNC Imaginative prose. Distribution of the two main types as a percentage of all.

All *to*-infinitives	All bare infinitives	All *to*-infinitives as % of total	All bare infinitives as % of total
611	654	48.3 %	51.7 %

Figure 10. BNC Imaginative prose. Distribution of all *to*-infinitives and all bare infinitives.



According to the data from the late 20th century the two main types of complements are used with almost equal frequency in the sub-corpus; in fact, the complements without the infinitival marker *to* were attested with a slightly higher rate of occurrence than their counterparts with the marker *to*.

For the quantitative analysis of the BNC Imaginative prose section data could be extracted with high precision and recall exploiting the metadata annotation of the corpus and the query languages provided by the *BNCweb* interface. As no detailed analysis is needed at this point, manually removing irrelevant tokens from the data is not exceedingly time-consuming. To maintain as large a sample size as possible, a sample of the sub-corpus was not taken for the quantitative analysis of different infinitival

complements; the numbers in Table 17 above include all the relevant tokens collected with the search heuristics presented in Section 2.5.2. For the semantic analysis, however, the trained eye of an analyst is needed. To escape the overwhelming task of carefully analyzing all 1265 tokens collected in the sub-corpus and dividing them into the 'aid' and 'contribute to' categories, a more manageable random sample of 25% was taken from all the collected tokens. Table 19 below presents the results of this analysis:

Table 19. BNC Imaginative prose. Complement/sense mapping for 25% of all collected tokens.

Complement	**Sense** 'aid'	**Sense** 'contribute to'
to-infinitive	14	46
NP + *to*-infinitive	93	1
bare infinitive	30	28
NP + bare infinitive	106	-
Total	**243**	**75**

As can be seen in Table 17 above, all four infinitival complement types were found after all the different inflected forms of *HELP*. The majority of the patterns with the *to*-infinitive were found to be instances of the 'contribute to' sense. (75)–(76) exemplify the 'aid' sense and examples (77)–(78) the 'contribute to' sense:

(75) Whatever happens, I'm going to stay with you and help to build up the practice till it's a going concern. (JYE 1853)
(76) The day before he left was a Saturday and Nick and Carrie were going to spend it helping to harvest the small hay field at Druid's Bottom. (EFJ 2189)
(77) Snow piled against it [a tent], threatening to destroy it, but this helped to protect it eventually as the snow wall hardened and compacted. (HTM 2564)
(78) Of course I've made him a pie — why shouldn't I? She won't. But if it helps to get them together I'll say she made it. (K8R 379)

In (75) and (76) there is an understood object that is helped to do something: in (75) *you* could be inserted after *help*, and in (76) some other people are helped in the harvest. In (77)–(78) the subject plays a part in the action described in the lower clause. These are closer to the 'contribute to' sense.

HELP + NP + *to*-infinitive patterns manifest the 'aid' sense with one exception in the material that went through the semantic analysis. (79)–(80) are examples of the 'aid' sense and (81) is the single instance deemed to fall into the 'contribute to' category.

(79) You can come and help me to carry some pictures, George dear . . . (ASE 998)
(80) I told you, Murphy, that although I am immensely grateful to you for helping me to repair the shed to make a hen house and for finding me the things I need, these are my hens and I shall look after them. (H8X 1856)
(81) . . . the Sutherlands made a fortune running the blockades with war supplies from the British and French, as well as helping the cotton export to flow from the ports of Wilmington, Charleston and Savannah. (FPF 753)

Again, (79)–(80) are quite clear cases of the 'aid' sense. In both the object NP refers to a human entity that can, indeed, be helped to do something. The object NP *the cotton export* in (81), however, is a non-human, non-reasoning entity which cannot be helped to do something in the 'aid' sense of the word. This, perhaps, gives rise to the 'contribute to' interpretation. The subject of the clause, *the Sutherlands*, had some sort of a role in *helping the cotton export to flow*; they contributed to it.

HELP + bare infinitive patterns were distributed quite evenly into the 'aid' and 'contribute to' classes. (82)–(83) exemplify the former and (84)–(85) the latter:

(82) 'Will you help carry the coffin?' he asked Preston before he left. (F9C 2765)
(83) More importantly, Emperor Henry had also agreed to send an army to help throw the Normans, once and for all, out of Apulia. (HRC 410)
(84) The skill, once learnt, helped make me a little more popular and outgoing at school, where there was a craze on for such things. (FR3 232)

(85) To help pay off the debts my mother got herself a job. (H7W 1706)

In (82)–(83) somebody or something is concretely aided in some task. In (84)–(85) there is, again, the sense of the subject playing a part in the action or promoting it.

Finally, *HELP* + NP + bare infinitive patterns were found to manifest the 'aid' sense:

(86) At least now, today, she could ease his suffering, help him forget Madeleine for a little while, make him see that he could still find happiness of a kind without her. (FS1 2482)
(87) He's been helping Mark put together the Business Plan and the presentation (AC2 913)

In all the tokens that went through the semantic analysis the object NP was either a personal pronoun or a proper name referring to a person or persons. This gave rise to the 'aid' interpretation.

Any categorical conclusions about the constraints imposed by the *horror aequi* principle in the BNC Imaginative prose are difficult to make at this point. Some sequences of *help* + bare infinitive are preceded by *to*, but others are not. A telling example that such constraints may still exist is recorded in (88):

(88) She did not beg her, as Cati did — Rosa had heard her — to help her be good, help her to be pure, and never have dirty thoughts . . . (GUX 613)

Instances of *to help* + *to*-infinitive do occur in the data, but in most cases where *help* is in the *to*-infinitive the following infinitive construction seems to a bare infinitive; a more exhaustive statistical analysis of possible *horror aequi* restrictions in present-day data will be presented in Section 6.3.

Insertions favoured infinitives with *to*:

(89) Above all, help us never to forget the teachings of the great sage. (FU8 766)

(90) Now his hands are holding my head, holding me in, helping me not to fly away, apart, dissolve. (CA3 2695)
(91) Francie, especially, she loved and would help her aunt gladly to mend his clothes. (FRC 2240)

In (89) *never* is inserted between the object NP and the infinitive. In (90) the insertion consists of the *not*-negation of the infinitive; as has been noted, negated infinitives seem to favour the marker *to*. In (91) the adverb *gladly* is inserted between the object NP and the infinitive. Altogether eight tokens with an insertion were found in the data and all of these were attested with the NP + *to*-infinitive complement.

Three passive constructions were collected, all with the marker *to*:

(92) He said they were helped by gypsies to get out of Romania and through Hungary. (HTJ 1711)
(93) . . . he could no longer hold a fork and had to be helped to eat. (GUK 3221)
(94) . . . the individual is guaranteed a much better chance to lead a life free from catastrophe and, if catastrophe occurs, a much better chance to be helped to recover. (HGS 1889)

For the study of argument structure and complementation, passives were related to their active versions. Based on these active versions they were placed into proper categories in Table 17 above, for example.

6. Findings

In this chapter I will move on from the data analysis of Chapter 5 to presenting the findings of this study. First, diachronic change in the distribution of infinitival complements will be charted. Secondly, I will discuss whether different patterns carry different senses of the verb and address the question whether *HELP* is undergoing grammaticalization. Finally, a statistical analysis of possible *horror aequi* constraints in present-day data is presented.

6.1. Diachronic change in the distribution of the infinitival complements of *HELP*

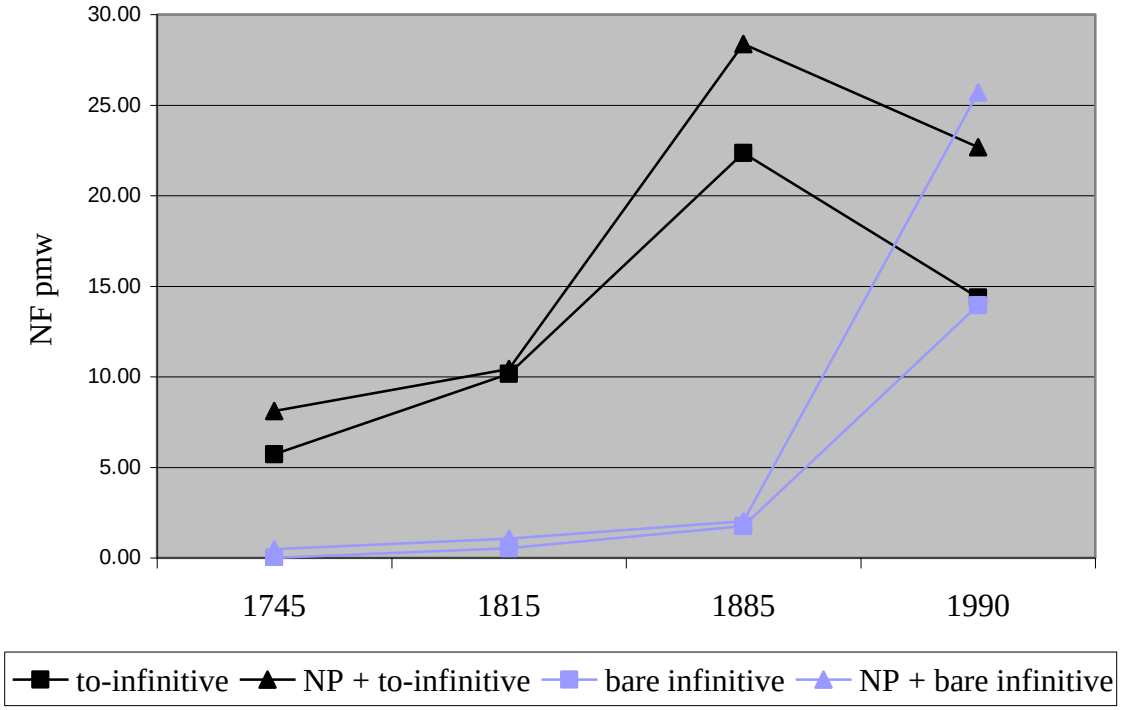
In this section the results obtained from the four sets of data presented in Chapter 5 are collapsed into one diachronic analysis. In particular, this section aims to provide an answer to the research aim i) presented at beginning of this thesis, which is reprinted here for convenience:

> i) Chart changes in the distribution of the infinitival complement patterns of the verb *HELP* from the early 18th century to present-day English.

As the research questions implies, the approach adopted in this section is mainly quantitative. The hypothesis that can be formed based on earlier studies on the subject (introduced above in Section 4.2) is that in BrE the complements with the *to*-infinitive were more used than complements with the bare infinitive until the second part of the 20th century. Data from the *OED* citation database suggest that instances of the bare infinitive after *HELP* were marginal from the 17th century until the dawn of the 20th century (Mair 2002, 123–124).

Figure 11 shows that there were significant changes in the frequency of occurrence of the four infinitival complements during the period under investigation:

Figure 11. Infinitival complements of *HELP* in the four sub-corpora, diachronic trends.[21]



As can be seen in Figure 11, the bare infinitive and NP + bare infinitive complements remain rare until the end of the 19th century in the genre under inspection here; only marginal increase in their usage was recorded. The difference to the present-day data is dramatic: the bare infinitive clause complements occur with almost equal frequency with the *to*-infinitive clause complements after *HELP*; the frequency of usage of the NP + bare infinitive complements has surpassed that of the NP + *to*-infinitive complements.

In Chapter 5 of this thesis the *to*-infinitive complements and NP + *to*-infinitive complements were put into one main category representing all infinitival complement

[21] The point of reference for each subsection of the CLMET was set at the middle point of the section. Hence for the CLMET 1, which covers the years 1710 to 1780, the point of reference was set at the year 1745, for example. The point of reference for the BNC was set at the year 1990. This convention will be used throughout the rest of this thesis.

clauses with the infinitival marker *to*; the same was done for the bare infinitive and NP + bare infinitive complements. For a diachronic perspective, the results of the comparisons of these two main types from all four sub-corpora are presented first as percentages of all infinitival complements in Figure 12, and then as normalized frequencies in Figure 13:

Figure 12. The two main types in the four sub-corpora as percentages of all infinitival complements, diachronic trends.

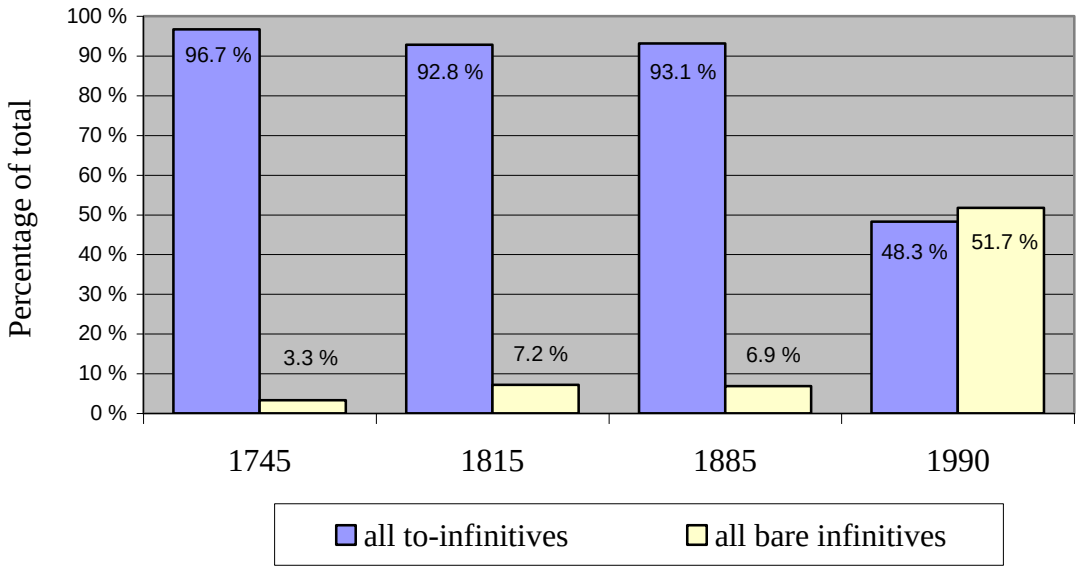
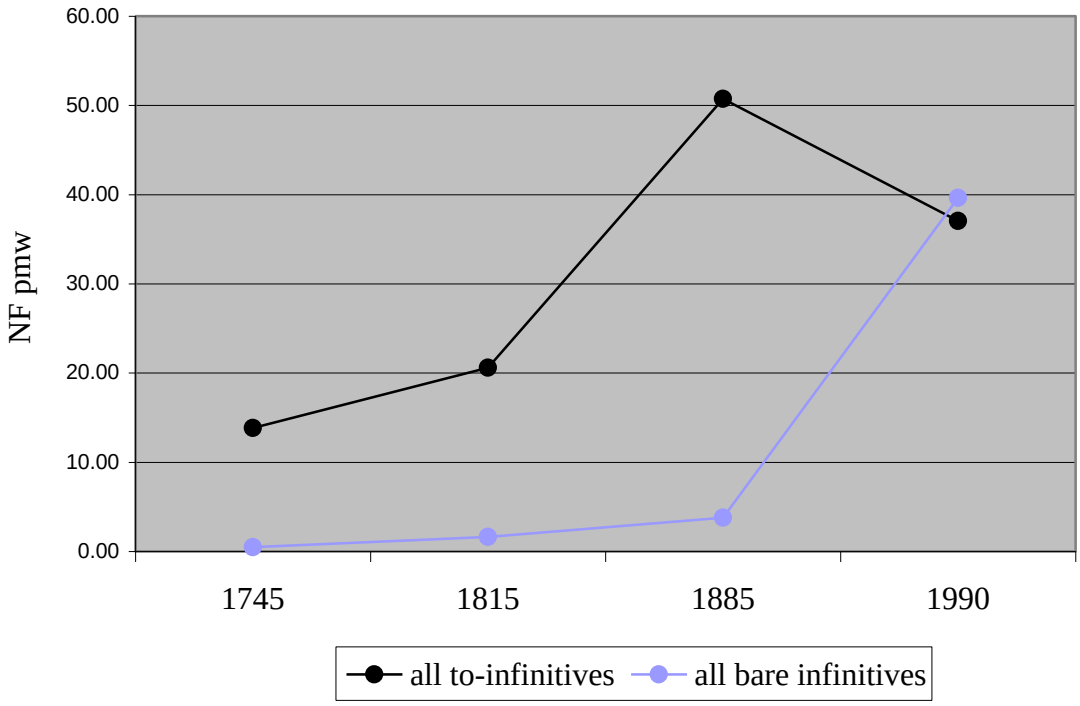


Figure 13. The two main types in the four sub-corpora as normalized frequencies, diachronic trends.



When the complements are collapsed into the two main types, the difference between historical and present-day data becomes even more telling: infinitival complements without the marker *to* remained somewhat marginal in the data from the 18th, 19th, and early 20th century; when we come to the present-day data, all bare infinitive constructions account for more than half of all the infinitival complements that *HELP* selects, and the relative frequency of the main type has risen from marginal to close to 40 instances pmw. As regards the other main type, infinitival complements with the marker *to*, there was a pronounced increase in its frequency of occurrence from the 17th century to the turn of the 20th century; when we come to the present-day data, however, its relative frequency has declined.

The question that begs to be answered is whether the decline from the CLMET 3 to the BNC in the usage of the patterns with *to* can be accounted by the even steeper

increase in the usage of the constructions without *to*; has the latter taken over some communicative functions from the former, or do they simply occur in free variation in present-day language?

6.2. Pattern/sense matching and grammaticalization

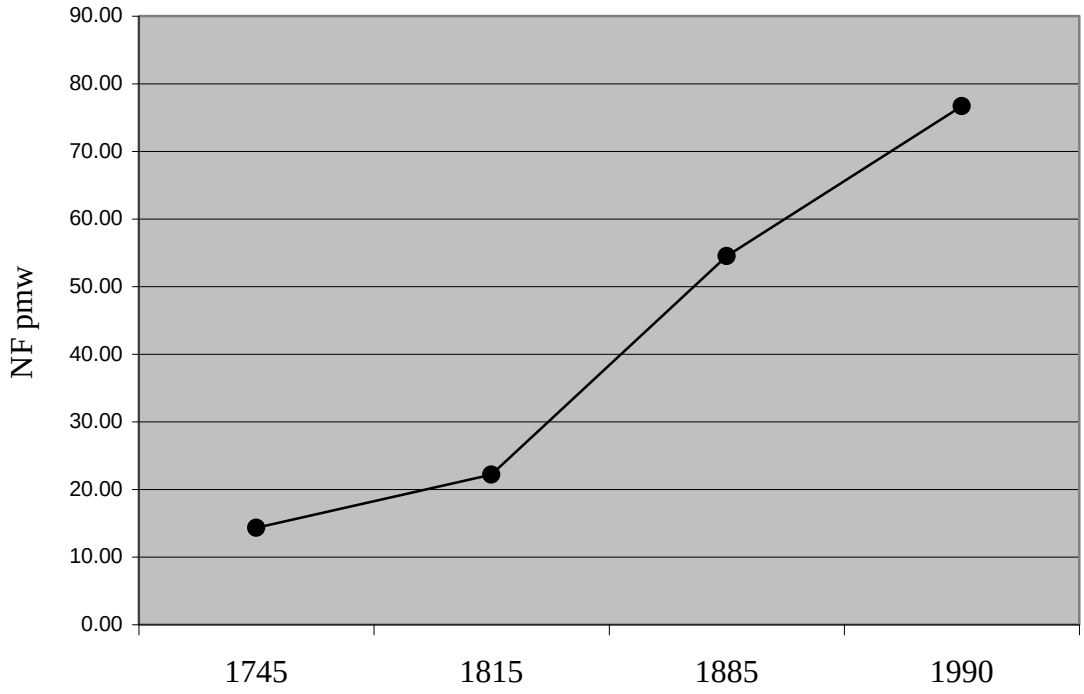
This section aims to provide answers to research questions ii) and iii) presented at the beginning of this thesis. These are reprinted here for convenience:

ii) Examine whether different patterns convey different senses of the verb.

iii) Address the question whether *HELP* is undergoing a process of grammaticalization.

Comparisons of the frequencies of usage between different infinitival complements among all infinitival complements miss one interesting aspect in diachronic development. As can be inferred from the data presented in Chapter 5, in addition to fluctuations within the group of all infinitival complements there has been a significant overall increase in the frequency of *HELP* + any infinitival complement:

Figure 14. All infinitival complements of *HELP* in the four sub-corpora, diachronic trend.



During the period investigated the overall frequency of *HELP* selecting an infinitival complement rose from 14.31 pmw in the CLMET 1 to a whopping 76.68 pmw in the BNC Imaginative prose. Figure 14 shows that the decrease in the frequency of the infinitival clause complements with the marker *to* after the CLMET 3 period is more than overlaid by the increase in the frequency of all bare infinitive constructions.

As was discussed in Section 4.3, some authors have attributed this steep overall increase to the process of grammaticalization. In any case, it is hard to see this drastic increase in any other way but as evidence that the meaning of *HELP* has widened; *HELP* is used in the language either to code cognitive contents which it did not code at all before, or to code contents which it did not code before as frequently as it does today. Though the semantic categories of ‘aid’ and ‘contribute to’ were motivated in Section 3.1 and used throughout Chapter 5, it has become evident that the semantics of different

patterns are far from being black-and-white; they are rather characterised by a million different shades of grey. Hence I will not try to provide an answer of the statistical-quantitative kind to the research questions on semantics; the approach in this section will rather be of qualitative nature.

HELP + *to*-infinitive constructions were relatively frequent in all sub-corpora. In all sets of data they were found to manifest the 'contribute to' sense in the majority of cases. The desemanticized 'contribute to' sense is hence not a new property of *HELP*; it was attested already in the 18th century with the pattern *HELP* + *to*-infinitive. This finding was surprising, as earlier literature placed the emergence of the 'contribute to' sense in the 20th century, and claimed that the rise of the bare infinitive complement is connected to the more grammaticalized sense. Yet in the data from the late 20th century instances of the pattern *HELP* + *to*-infinitive were found to manifest the 'contribute to' pattern in the majority of tokens.

HELP + bare infinitive constructions were rare in all CLMET sub-corpora. There were none, two, and seven tokens of this type in the CLMET parts 1, 2, and 3, respectively. Whenever they were attested in the historical data, they were found to fall into the simplified sense category of 'aid'. In the majority of cases they occurred in environments where *HELP* itself was in the *to*-infinitive; it appears to be the case that in the historical data studied in this thesis the pattern *HELP* + bare infinitive occurs mainly for syntactic reasons and that the pattern has no specialized meaning. The situation in the present-day data is different: the pattern is in much more frequent use and its occurrences are divided quite evenly into the 'aid' and 'contribute to' categories. Hence the pronounced increase in the rate of occurrence of this pattern can be partly accounted by

the widening of the meaning of the pattern *HELP* + bare infinitive; it is, indeed, true that in about half of its occurrences in present-day data the pattern manifests the desemanticized 'contribute to' sense. Though the issue clearly begs further investigation, this increase can be seen as contributing factor to the decrease of the *HELP* + *to*-infinitive pattern from the CLMET 3 to the BNC; whereas in the historical data the 'contribute to' sense was recorded solely with the *to*-infinitive, in present-day data this communicative function may have partly been taken over by the bare infinitive construction.

With the two complement patterns that have the intervening NP between the matrix and the infinitive, the semantic analysis may quickly become circular: many tokens might be assigned a reading from the grey area between the clear-cut 'aid' and 'contribute to' categories, and the analyst may be left with the absence or presence of the object NP with a human reference as the only clue towards placing it into one of the two simplified sense categories. In example (95) below, for instance, the meaning of *helped* is quite causative, and it could be, perhaps, paraphrased with *make* in the sense of 'contribute to' or 'play a part in' with little change to meaning:

(95) My reaction finally helped Matthew to understand how much I love him. (JY6 3930)

The paraphrase would read as *My reaction finally made Matthew understand how much I love him*. In the vast majority of tokens the object NP has a human reference that can be helped to do something; this property of the NP was a deciding factor in placing tokens such as (95) into the 'aid' category. The few examples where the object NP was a non-human, non-reasoning entity, which cannot be helped to do something in the 'aid' sense of the verb, were assigned the 'contribute to' reading. One such token was presented in (81), reprinted here as (96):

(96) . . . the Sutherlands made a fortune running the blockades with war supplies from the British and French, as well as helping the cotton export to flow from the ports of Wilmington, Charleston and Savannah. (FPF 753)

Though the main focus of this thesis is not on semantic roles, it appears to the present author that the arguments that *HELP* subcategorizes for tend to have different semantic roles with different senses of the verb. With the 'aid' sense, the object NP of the patterns denotes a human entity, whereas with 'contribute to' sense it denotes a non-human entity.

My results show some support for the claims that *HELP* is undergoing a process of grammaticalization. The patterns without the object NP are used with the more desemanticized sense of the verb quite frequently. Because of the nature of the object NP in those patterns that have it, these patterns are assigned an interpretation closer to the prototypical 'A helps B to do C' sense. My findings do not support the claims that in present-day language the bare infinitive complements are more likely to code the desemanticized meaning than the *to*-infinitive complements. The small decline in the frequency of the *to*-infinitive complements in the 20th century can be, however, accounted for by the even steeper increase in the frequency of the bare infinitive complements; the bare infinitive complements might be taking over the communicative function of the desemanticized 'contribute to' sense from the *to*-infinitive complements. It remains to be seen whether this development continues.

6.3. *Horror aequi*: *help*/*to help* difference[22]

The goal of this section is to shed light on the research aim number iv) listed in the Introduction, which is reprinted here for convenience:

> iv) Discuss different contextual factors that may condition the choice between competing complements. These factors may be syntactic, semantic, or extra-linguistic. In the data analysis special emphasis will be on possible effects caused by the phenomenon known as *horror aequi*.

To study the constraints that the *horror aequi* principle might have on complement selection I decided to compare all instances of the verb form *help* not preceded by *to* and all instances of *to help* in the BNC Imaginative prose. The approach in this section is hence not diachronic, but provides a synchronic glimpse into possible restrictions in BrE imaginative writing of the late 20th century. Table 20 notes the frequencies of different infinitival complements preceded by *help* and *to help*:

Table 20. Infinitival complements followed by *help* and *to help* in BNC Imaginative prose.

	Infinitives with *to*		Infinitives without *to*	
	to-infinitive	NP + *to*-infinitive	bare infinitive	NP + bare infinitive
help	70	147	54	152
to help	5	29	114	121

In Section 4.5 I hypothesized that an intervening NP between the matrix verb and the infinitive clause potentially cancels out *horror aequi* effects. Let us first consider complements without the intervening object NP. *Help* without the preceding *to* is followed 70 times by the *to*-infinitive complement and 54 times by the bare infinitive complement. The sequences of *to help* are followed five times by the *to*-infinitive

[22] Throughout this subsection *help* is used to refer to instances where *help* is not preceded by *to*. *To help* refers to instances where *help* itself occurs in the *to*-infinitive.

complement and 114 times by the bare infinitive complement. These numbers form a 2 x 2 table, to which the chi-square test can be applied:

Table 21. 2 x 2 table for the chi-square test. Complements without the object NP.

	to-infinitive	bare infinitive
help	70	54
to help	5	114

For these numbers the chi-square test gives the X^2 score of 75.26, which is well above the threshold at the level $p < .001$ and which marks the difference as of very high significance.[23]

When it comes to the complements with the object NP, *help* is followed 147 times by the NP + *to*-infinitive complement and 152 times by the NP + bare infinitive complement. The sequences of *to help* are followed 29 times by the NP + *to*-infinitive complement and 121 times by the NP + bare infinitive complement. The 2 x 2 table formed for the complements with the object NP is presented below:

Table 22. 2 x 2 table for the chi-square test. Complements with the object NP.

	NP + *to*-infinitive	NP + bare infinitive
help	147	152
to help	29	121

For these numbers the X^2 value is 36.05, which is significant at the level $p < .001$; again, the effect that the *help*/*to help* difference has is of very high statistical significance.

Based on these tests we can safely conclude that the effect that *horror aequi* has on complement selection when the *help*/*to help* difference is contrasted is highly significant. The hypothesis that an intervening object NP might cancel out *horror aequi* effects can be refuted for the most part: the effect that the *help*/*to help* difference has on complement

[23] The critical value for the significance level $p < .001$ is 10.827.

selection as regards the complements with the object NP is statistically highly significant, too.[24] [25]

[24] It should be noted, however, that the X^2 score of 75.26 for the complements without the object NP means higher statistical significance than the X^2 score of 36.05 for the complements with the object NP. Nonetheless, both are well above the critical value of 10.827 for the significance level $p < .001$, which is considered to be the threshold for very high significance.

[25] As an anonymous reader of an earlier version of this thesis correctly pointed out, there are several other factors in addition to the *horror aequi* principle that might favour the use of one complement over the other, and hence have an impact on the numbers in Tables 20–22; these factors were discussed in the subsections of Chapter 4. I take the point, but would still maintain that the numbers cited in Tables 20–22 are legitimate and that extraction and insertion environments (that potentially favour infinitives with *to*), for instance, need not be excluded from the calculations; if extraction and insertion environments were to be excluded from the numbers in Tables 20–22, it would be difficult to justify why tokens that can be seen as affected by the distance principle (in a way that potentially favours infinitives without *to*), for example, were not to be excluded. The factors bearing on the complement selection of *HELP* present a complex interplay of factors, and I argue that the influence of one factor cannot be excluded in the presence of another. However, the numbers in Tables 20–22 do not include passive constructions; because passive transformations were related to their active versions, it is not, perhaps, appropriate to consider these in Tables 20–22, as the active paraphrases are not found directly in the primary data.

7. Conclusion

In this thesis I have analyzed 1595 authentic instances of *HELP* controlling an infinitival complement clause. These tokens were collected from corpora of both historical and present-day British English representing written prose language. With the help of this set of data, several interesting results were obtained.

Diachronic change in the distribution of the infinitival complement patterns during the studied period was dramatic: from the 18th to early 20th century the frequency of the infinitives with the marker *to* clearly rose, but during the 20th century their frequency declined. The bare infinitive and NP + bare infinitive complements were marginal in all sets of historical data, but their frequency mushroomed during the 20th century, and nowadays *HELP* is more likely to choose an infinitive without the infinitival marker *to* in British English prose.

Furthermore, the overall frequency of *HELP* selecting an infinitival complement has risen drastically, which can partly be explained by the desemanticization of the meaning of *HELP*. The *HELP* + *to*-infinitive pattern was found to manifest the 'contribute to' sense in the majority of tokens in all sets of data. The fact that this sense was found already in the historical data and especially with the *HELP* + *to*-infinitive pattern was a surprising discovery; some earlier studies on the matter linked the grammaticalized 'contribute to' sense to the rise of the bare infinitive, which took place in the 20th century. In the historical data, the bare infinitive complements were found to occur with the prototypical 'aid' sense of the verb, which, again, was not expected; in the present-day data, however, the pattern occurs with the desemanticized 'contribute to' sense in about half of its occurrences. Furthermore, it was noted that in the historical data

the use of the bare infinitive complement could often be accounted for by syntactic factors (*horror aequi*). The rise of the *HELP* + bare infinitive pattern with the 'contribute to' sense coupled with the decline of the *HELP* + *to*-infinitive pattern in the 20th century was treated as a sign that the former might be taking over the communicative function of the desemanticized 'contribute to' sense from the latter. However, even in data from the BNC, the 'contribute to' sense was found to occur more often with the *to*-infinitive than with the bare infinitive. Patterns with the object NP were deemed to fall into the 'aid' category almost exclusively.

Several contextual factors that may condition the choice between the complements were discussed, and special emphasis was placed on the effect that the *horror aequi* principle has on complement selection; indeed, the impact that *horror aequi* has was proven to be statistically highly significant. This was true irrespective whether an object NP occurred in the complement pattern or not.

With these results, I have fulfilled the aims that I set for this study at the outset. Yet there remain several interesting aspects of the infinitival complementation of the verb *HELP* that this thesis only scratched the surface of, or that were altogether beyond the scope of it. These include, for instance, the impact of different inflected forms of *HELP* on complement selection and the possible effect of the length of the intervening NP. Furthermore, the influence of cognitively complex environments, such as insertion and extraction environments, was only touched on in this study. These issues and many others, however, will be left for future endeavours.

References

Primary sources of data

The Corpus of Late Modern English Texts (CLMET)

The British National Corpus (BNC)

Secondary sources

Anderson, John, ed. 1982. *Language Form and Linguistic Variation: Papers Dedicated to Angus McIntosh*. Amsterdam: John Benjamins B.V.

Arts, Bas and Charles F. Meyer, eds. 1995. *The Verb in Contemporary English: Theory and Description*. Cambridge: Cambridge University Press.

Ball, C. N. 1994. "Automated text analysis: cautionary tales." *Literary and Linguistic Computing* 9, 4: 295–302.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Bolinger, Dwight. 1968. "Entailment and the meaning of structures." *Glossa* 2, 2: 119–127.

Burnard, Lou, ed. 2007. Reference Guide for the British National Corpus (XML Edition). [Internet] Oxford, Oxford University. Available at http://www.natcorp.ox.ac.uk/docs/URG/. [Accessed 8 February 2011]

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press.

Dalton-Puffer, Christiane, Dieter Kastovsky, Nikolaus Ritt, and Herbert Schendl, eds. 2006. *Syntax, Style and Grammatical Norms*. Bern: Peter Lang.

Davies, William and Stanley Dubinsky. 2004. *The Grammar of Raising and Control*. Oxford: Blackwell.

De Smet, Hendrik. 2005. "A corpus of Late Modern English texts." ICAME Journal 29, 29: 69–82.

Dixon, Robert M. W. 1991. *A New Approach to English Grammar, on Semantic Principles*. Oxford: Oxford University Press.

Dixon, Robert M. W. 2005. *A Semantic Approach to English Grammar*. Oxford: Oxford University Press.

Haegeman, Liliane. 1991. *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Heine, Bernd, Ulrike Claudi, and Friederike Hünnemeyer. 1991. "From cognition to grammar – evidence from African languages." In *Approaches to Grammaticalization*, vol. 1, eds. Elizabeth Traugott and Bernd Heine, 149–187. Amsterdam: John Benjamins.

Hoffmann, Sebastian, Stefan Evert, Nicholas Smith, David Lee, and Ylva Berglund Prytz. 2008. *Corpus Linguistics with BNCweb – a Practical Guide*. Frankfurt am Main: Peter Lang.

Hunston, Susan and Gill Francis. 2000. *Pattern Grammar: A Corpus Driven Approach to the Lexical Grammar of English*. Amsterdam. John Benjamins Publishing Company.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.

Leech, Geoffrey. 1968. "Some assumptions in the metatheory of linguistics." *Linguistics* 6, 39: 87–102.

Leech, Geoffrey and Jan Svartvik. 2002. *A Communicative Grammar of English*. Harlow: Pearson Education Limited.

Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith. 2009. *Change in Contemporary English*. Cambridge: Cambridge University Press.

Lind, Åge. 1983. "The variant forms *help to*/*help* ∅." *English Studies* 64, 3: 263–275.

Linqvist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

Mair, Christian. 1995. "Changing patterns of complementation, and concomitant grammaticalization, of the verb *help* in present-day British English." In *The Verb in Contemporary English: Theory and Description*, eds. Bas Arts and Charles F. Meyer, 258–272. Cambridge: Cambridge University Press.

Mair, Christian. 2002. "Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora." *English Language and Linguistics* 6, 1:105–131.

McEnery, Anthony and Zhonghua Xiao. 2005. "*HELP* or *HELP to*: What Do Corpora Have to Say?" *English Studies* 86, 2: 161–187.

McEnery, Tony and Andrew Wilson. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

The Oxford English Dictionary. 2nd ed. 1989. OED online. [Internet] Oxford: Oxford University Press. Available at http://www.oed.com. [Accessed January 2011]

Perlmutter, David and Scott Soames. 1979. *Syntactic Argumentation and the Structure of English*. Berkley: University of California Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Quirk, Randolph and Jan Rusiencki. 1982. "Grammatical data by elicitation." In *Language Form and Linguistic Variation: Papers Dedicated to Angus McIntosh*, ed. John Anderson, 379–394. Amsterdam: John Benjamins B.V.

Rohdenburg, Günter. 1996. "Cognitive complexity and increased grammatical explicitness in English." *Cognitive Linguistics* 7, 2: 149–182.

Rohdenburg, Günter. 2003. "Cognitive complexity and *horror aequi* as factors determining the use of interrogative clause linkers." In *Determinants of Grammatical Variation in English*, eds. Günter Rohdenburg and Britta Mondorf, 205–249. Berlin: Mouton de Gruyter.

Rohdenburg, Günter. 2006. "The role of functional constraints in the evolution of the English complementation system." In *Syntax Style and Grammatical Norms*, eds. Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt, and Herbert Schendl, 143–166. Bern: Peter Lang.

Rohdenburg, Günter and Britta Mondorf, eds. 2003. *Determinants of Grammatical Variation in English*. Berlin: Mouton de Gruyter.

Rudanko, Juhani. 1996. *Prepositions and Complement Clauses : A Syntactic and Semantic Study of Verbs Governing Prepositions and Complement Clauses in Present-Day English*. New York: State University of New York Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, John. 1996. EAGLES. Prelimininary recommendations on Corpus Typology. [Internet] Available at http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html. [Accessed 20 January 2011]

Traugott, Elizabeth and Bernd Heine, eds. 1991. *Approaches to Grammaticalization*. 2 vols. Amsterdam: John Benjamins.

Traugott, Elizabeth and Ekkehard König. 1991. "The semantics-pragmatics of grammaticalization revisited." In *Approaches to Grammaticalization*, vol. 1, eds. Elizabeth Traugott and Bernd Heine, 189–218. Amsterdam: John Benjamins.