

Jill Bowie and Sean Wallis

2 The *to*-infinitival perfect: A study of decline

Abstract: The English *to*-infinitival perfect (as in *She claims to have seen him*) has not received the same attention as the present perfect. In this paper we examine its changing use in written American English over the last 200 years, using data from the *Corpus of Historical American English* (COHA). This reveals a reduction of about 80 % over this period, against a baseline of other past-referring forms (present perfect and past tense verbs). Secondly, we examine contexts of decline, focusing on the most frequent verb collocates of the *to*-infinitival perfect in COHA (such as CLAIM in the example above) on the premise that these collocates identify the semantic contexts in which the *to*-infinitival perfect may be used. Collocates are divided into subgroups based on semantic and grammatical criteria, including possible alternation patterns to the *to*-infinitival perfect. This procedure exposes a rich variation in the behaviour of both subgroups and individual verbs.

1 Introduction

Many studies in English linguistics have focused on the present perfect, whereas other types of perfect have received less attention. In this study we look at the *to*-infinitival perfect, as in *A witness is said to have seen the whole event*, considering the changing usage of this structure in written American English (AmE) over the last 200 years.

We first need to set this structure in context as one subtype of the English perfect. In all perfect subtypes, the perfect auxiliary HAVE is followed by a verb in the past participle form. The auxiliary occurs in the present tense in the present perfect (as in *She has spoken to them*), but it can also occur in the past tense, giving the past perfect (*She had spoken to them*). Additionally, it can occur in non-finite forms, within gerund-participial clauses (*Having spoken to them, she went home*), bare-infinitival clauses (*She must have spoken to them*), or *to*-infinitival clauses (*She seems to have spoken to them*).

All subtypes of the English perfect typically function to express anteriority (i.e. “pastness” relative to a reference point). This supports a treatment of the

Jill Bowie and Sean Wallis, University College London

perfect as an analytically expressed tense system (e.g. Huddleston and Pullum 2002). However, some grammarians analyse the perfect in terms of aspect (e.g. Quirk et al. 1985), and the correct analysis has been much debated, largely because of the added semantic complexity associated with the present perfect (see, e.g., Comrie 1976; Elsness 1997; Werner 2014; Werner, Seoane, and Suárez-Gómez, this volume). The present perfect generally presents a situation as occurring within (or even continuing through) a time span beginning in the past and leading up to the present, with an added dimension of “current relevance” (i.e. a focus on the present repercussions of the situation). This specialised use contrasts with that of the inflectionally marked past tense (used to represent a time as wholly in the past rather than connected to the present; see the contributions in this volume by Fuchs; Fuchs, Götz, and Werner; Seoane; Van Rooy).

This contrast, however, is neutralised in the other types of perfect, which can correspond to either a present perfect or a simple past (Huddleston and Pullum 2002: 146). For instance, compare (1) and (2).

- (1) *The government claims to have achieved great things since the election.*
- (2) *The neighbour claims to have heard three gunshots shortly after midnight.*

The content of the claim in (1) would be rendered as “We have achieved great things since the election” (with a present perfect), while that of (2) would be “I heard three gunshots shortly after midnight” (with a simple past).

A number of corpus-based studies have looked at variation and change in the present perfect (including, among others, Elsness 1997, 2014; Hundt and Smith 2009; Yao and Collins 2012; Bowie, Wallis, and Aarts 2013a). Most of the evidence indicates that its frequency has declined since 1800 in AmE, and some of the evidence also suggests a decline in British English (BrE) – though this is less clear – while for the present day there is general agreement that the structure is more frequent in BrE than in AmE.

There has been more limited corpus investigation of other (non-present) types of perfect, which are of much lower frequency than the present perfect. The results show some parallels with the findings for the present perfect, in terms of American versus British contrasts and of trends of decline over time. Elsness (1997) reports written data showing a higher frequency of the past and infinitival perfect forms in BrE than in AmE of the present day, and a decline in the fre-

quency of these forms in AmE since 1800.¹ Yao and Collins (2013) report written data showing a higher frequency of all non-present perfect types in BrE than in AmE of both the 1960s and 1990s, and a decline of the past perfect in both varieties from the 1960s to the 1990s. In our own previous research (Bowie and Aarts 2012; Bowie, Wallis, and Aarts 2013a), we investigated change over recent time (1960s to 1990s) in all forms of the perfect in spoken BrE in the *Diachronic Corpus of Present-Day Spoken English* (DCPSE), finding a decline over this period in the past and infinitival perfects. Thus, the research to date on non-present perfect structures shows that they are worthy of further investigation in terms of variation and change.

Our focus in the present study is the *to*-infinitival perfect. For this form, our previous research on spoken BrE from the 1960s to 1990s found a steep decline ($52\% \pm 26\%$)² in frequency relative to that of *to*-infinitives in general (Bowie and Aarts 2012). This decline presents a puzzle. It is not in general possible to simply omit the perfect without changing the meaning: there is a clear temporal contrast between *He is believed to be in Paris* and *He is believed to have been in Paris*, since in the latter the perfect expresses anteriority to the present time referred to by *is*. If people are using fewer *to*-infinitival perfects, how might they express the meanings conveyed by this form?

As the *to*-infinitival perfect is a relatively infrequent form (with only 127 occurrences in DCPSE, a corpus of c. 885,000 words), we need a larger corpus in order to investigate it in more detail. It is also of interest to look at longer-term trends in the use of this form. For the present study we use the 400-million-word *Corpus of Historical American English* (COHA; Davies 2012), which covers the period from 1810 to 2009 and includes text from the following genres: fiction, newspaper, magazine, and a fourth category, “other non-fiction”. Our research questions are as follows:

- Do we see a decreasing inclination to use the *to*-infinitival perfect in written AmE over the last 200 years?
- What are the most frequent structural and/or collocational contexts for this construction? Are there alternants for these which express a similar meaning? If so, are these alternants increasing in frequency at the expense of the construction we are investigating?

¹ For the infinitival perfect forms, we performed our own calculations based on combinations of the figures provided by Elsness for several different constructions involving either the bare infinitival perfect or the *to*-infinitival perfect (1997: 104, 267–268).

² Errors are estimated using the Wilson score interval at a 0.05 error rate, i.e. predicting that it is 95% probable that the true “population” value is within this range (Wallis 2013).

Our study is in three stages. First, we investigate frequency variation of the *to*-infinitival perfect in COHA over time. We discuss different ways of measuring rates of change, going beyond the commonly used baseline of words (typically, per-million-word adjusted frequency measures) to more linguistically relevant baselines which allow us to test for the robustness of the patterns we find. In previous studies using DCPSE and its software, ICECUP (Nelson, Wallis, and Aarts 2002), we were able to exploit the parse annotation to construct precise queries for the studied perfect form and a baseline category (such as tensed past-marked VPs; see Bowie, Wallis, and Aarts 2013a: 328). Although COHA is not a parsed corpus, the part-of-speech tagging and flexible search options offered by the web interface do permit a careful researcher to devise search query strings which approximate to constructional queries.

Second, we wish to identify the limited set of contexts within which the *to*-infinitival perfect occurs, since it is in these contexts that replacement with alternate forms is possible. To help us achieve this, we draw on data from a smaller, parsed corpus which allows us to carry out precisely specified searches for grammatical structures. The corpus used is ICE-GB, the British component of the *International Corpus of English* (Nelson, Wallis, and Aarts 2002).

Finally, we return to COHA to examine these contexts, and where feasible, enumerate plausible grammatical and lexical alternation patterns. For instance, we find that in COHA forms of the verb *SEEM* are the most frequent left collocates of the *to*-infinitival perfect. Enumerating the limited number of verbs found in this position, we examine potential alternation patterns such as (3) versus (4).

(3) *He seems to have known about it.*

(4) *It seems that he knew about it.*

In (3) we find a *to*-infinitival perfect, while in (4) *seems* takes as its complement a finite clause instead of an infinitival clause (and has impersonal *it* as subject). An obvious lexical alternant to be considered here is *APPEAR*, which is semantically close to *SEEM* and grammatically able to replace it in structures like (3) and (4). In fact there is a wide range of alternative expressions that speakers might use to express reasonably similar shades of meaning. These may differ more radically from (3) in terms of grammatical structure and/or lexical content, so that there is no clear-cut limit: *I suspect he knew about it*; *The evidence suggests he knew about it*; *He apparently knew about it*; and so on. It is not feasible in this study to investigate the full range of possible alternative expressions for every context in which a *to*-infinitival perfect occurs, and we therefore focus on a selection of closer structural and lexical alternants.

The plan for the paper is as follows. In Section 2 we introduce COHA and report on frequency variation of the *to*-infinitival perfect in this corpus over time. Section 3 describes ICE-GB and presents a frequency profile for the various grammatical contexts in which the *to*-infinitival perfect occurs in this corpus. We then go on to examine the most frequent contexts in COHA in Section 4, focusing on verbal left collocates, and considering some possible alternation patterns. Our conclusion is given in Section 5.

2 The decline of the *to*-infinitival perfect in COHA

This section briefly introduces COHA (2.1) before reporting on frequency variation of the *to*-infinitival perfect in this corpus over time (2.2).

2.1 COHA

The *Corpus of Historical American English* (COHA) is a part-of-speech tagged megacorporus of written English. It contains over 400 million words of written AmE sampled for every year from 1810 to 2009, with approximately 2 million words per year. It was released in late 2010 and is freely available over a web interface (<http://corpus.byu.edu/coha>).³ Davies (2012) provides an overview of the corpus.

COHA comprises four main text genres. Figure 1 depicts the proportion of words collected for each of these genres over time (see Appendix A for the raw word frequencies).⁴ It can be seen that the sampling is not fully consistent over time. Fiction comprises approximately half the material over the decades, and magazines contribute about a quarter from the 1820s onwards. Newspapers and non-fiction provide the remaining quarter, but their proportions are less evenly distributed over the earlier decades. Newspapers are included only from the 1860s onwards, and their share increases in later decades. The total quantity of material sampled per decade is also somewhat variable, with less than 20 million words for some of the early decades.

When we performed our analyses, we found that the data for the 1810s, much smaller in size in any case (at 1.18 million words), obtained observations that were

³ More recently, in March 2015, the corpus has also become available to purchase for download. For the research reported here, we used the old web interface, currently at <http://corpus.byu.edu/coha/old>.

⁴ The data are based on raw word frequencies under “Composition of the corpus”.

inconsistent with the remainder of the data – a fact that was true across many different analyses. The simplest explanation for this can be seen in the graph: the sampling for this decade is quite distinctive and this fact may skew the results. For this reason we decided to omit the first decade from our analyses.

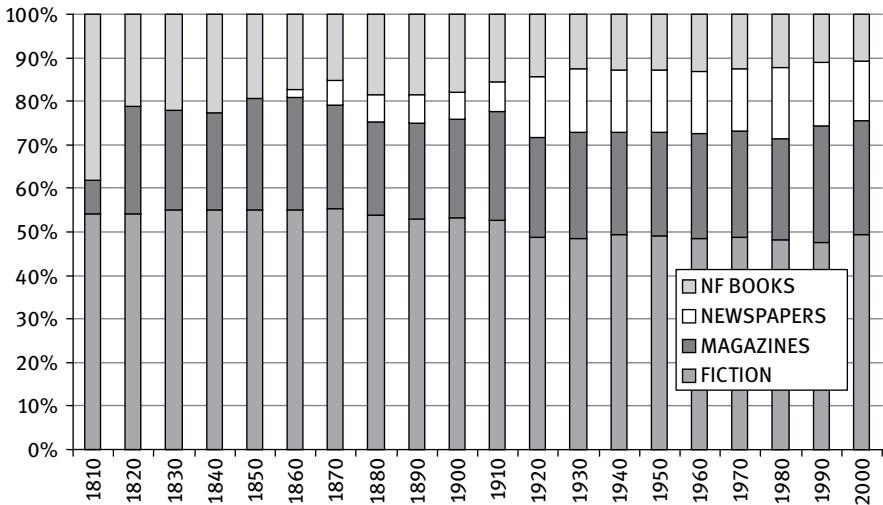


Figure 1: Distribution of text samples over the decades in COHA.

At the time of writing (2015), the web interface provides four different search and display options. To study frequency variation, we used the “Chart” option. This outputs a frequency breakdown by decade, providing both raw and per-million-word figures. From this display, the researcher can select any decade to see examples of the search item within their sentence context. COHA is annotated for parts-of-speech (POS), using the C7 tagset of the CLAWS tagger.⁵ Search options include specifying words, lemmas and POS tags, for all of which a simple wildcard system can be employed. As well as the usual lexical wildcard, applying a wildcard to a POS tag supports simple grammatical abstraction, such as extracting all verbs irrespective of any other coded features.

⁵ For the list of tags, see <http://ucrel.lancs.ac.uk/claws7tags.html>.

2.2 Frequency of the *to*-infinitival perfect over time

To search for the *to*-infinitival perfect we used the search string “to have [v?n*]”, i.e. the sequence (*to* + *have* + past participle verb). The third element in the string matches all items tagged as past participle verb forms. Note that this will not identify every example of the *to*-infinitival perfect: any string including one or more intervening words will not be found, so *to have started* will be included, whereas *to have already started* or *to already have started* will not. This represents an underestimate of approximately 2–3 % of genuine examples.⁶

Figure 2 plots the results of this search ($n = 66,219$) as a simple proportion per word (the web interface per-million-word graph produces a similar picture, without error bars). This graph indicates an almost linear decline in the overall use of the form (an “exposure rate”) over the period. Note that the expected distribution of a simple decline would be an “S” curve (a “logistic curve”). A linear decline in probability is steeper than this, and implies that the baseline is less than optimal.

This graph does not simply describe this sample: we can establish that the observed trend represents a genuine decline in AmE. Each point includes an “I”-shaped error bar, which identifies the region where the actual proportion in the population of comparable American written English is most likely to be found. These error bars are computed using the Wilson score interval method (Wallis 2013) with a 95 % range. This means we are 95 % certain that the true population value is within this range.⁷

To “read” the intervals in these graphs, note that where intervals do not overlap, the difference between observations is guaranteed to be statistically significant, i.e. a resample from the population will find a difference 19 times out

⁶ When sequences with an intervening word in either position are examined, it can be seen that the great majority are either not genuine examples of the *to*-infinitival perfect (e.g. *to have it known*), or involve instances with an additional past participle (perfect + passive, e.g. *to have been loved*) which are also included in the search string without intervening words. If only sequences with intervening words tagged as adverbs are considered, then most examples are genuine. Such sequences would represent 2 % of a corrected total that included them. Longer intervening sequences (such as multiple adverbs or preposition phrases like *in fact*, *at that stage*) also occur, but string searches show them to be much less frequent. We can compare results from ICE-GB (see Section 3), a parsed corpus, where it is possible to identify all examples of the required structure regardless of intervening words: in that corpus, only 1 % of instances of the *to*-infinitival perfect (2 out of 192) involve intervening words (a single adverb in each instance).

⁷ An anonymous reviewer asks whether this takes account of the missing 2–3 % of cases. The answer is no: it cannot do so, as it is a statistical measure based on the data we actually used (the cases without intervening words).

of 20. Slightly overlapping intervals *may* represent a significant difference, but a separate statistical test (the Newcombe-Wilson test) is required to determine this. Any pair of intervals where one point falls in the other's interval range should be considered effectively constant (i.e. the difference is non-significant). Thus in Figure 2, we cannot claim that the rate falls between 1840 and 1850, or that it rises again in the 1860s. For more information see Wallis (2013).

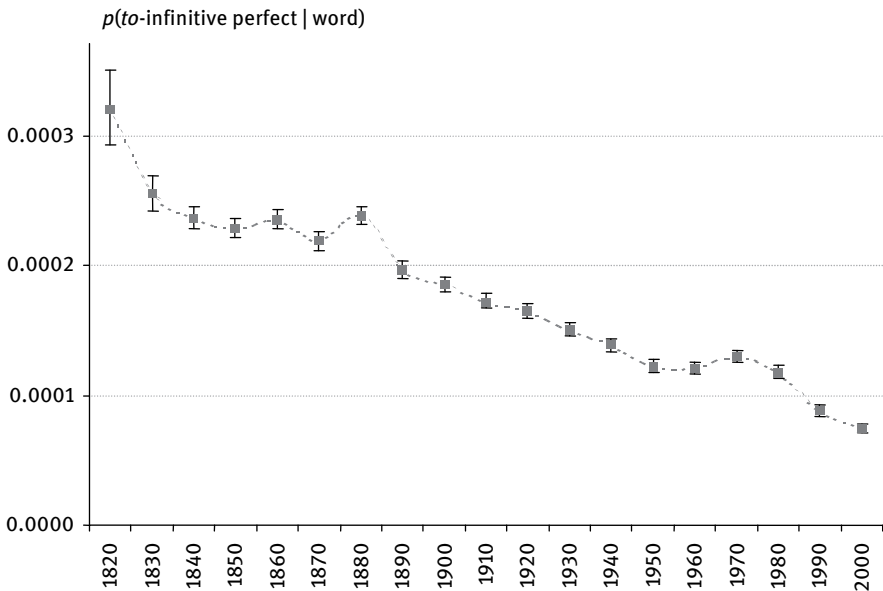


Figure 2: The *to*-infinitival perfect measured per word in COHA, by decade, with 95 % Wilson score intervals for each observation representing the likely range in the population.

Although per-million-word measures are commonly used in corpus linguistic studies, it is usually preferable to use other baselines. When any studied form is varying, the *opportunity* to use that form will almost certainly also vary.⁸ To obtain reliable results we must first factor out variation due to opportunity in

⁸ For example, Bowie, Wallis, and Aarts (2013b) show that the frequency of tensed verb phrases per word in the DCPSE corpus varies over genre and time. Suppose we were to study a subtype of tensed verb phrase which always had a constant 20 % share of cases. The correct conclusion would be that the proportion was stable over time and genre. But were we to use a per-million-word baseline to study this subtype, we would instead see the original undulating and varied pattern!

order to focus on the *choices* that speakers and writers make. Conflating these two types of variation, in the opportunity to choose and in the choices made, has a further undesirable consequence: it makes the reproduction of results between corpora more difficult than otherwise.

In our case we are interested in the grammatical choices made by writers, rather than their overall authorial choices, or, potentially, the effect of external constraints. We wish to investigate whether US writers have decreasingly chosen to use the *to*-infinitival perfect when the opportunity arises. We therefore address our hypothesis by examining contexts where the option to use this form arose in the first place, i.e. either (a) by setting observed frequencies against those of a set of alternants, or (b) by comparing them against the observed frequency distribution of a meaningful superordinate category. In this section, we employ method (b), as we are concerned with the overall data and have not yet considered the range of contexts within which the structure occurs.

We first consider a baseline of *to*-infinitives in general, since this represents the structural context within which the form in question can occur (the *to*-infinitival perfect is, logically, a subset of *to*-infinitives). This eliminates variation that may be due to changes in the use of *to*-infinitival complementation. The non-finite complementation system in English is still undergoing change. Mair (2006: 215) notes that “there is every indication that usage of infinitives, participles and gerunds has continued changing with unabated vigour [...] in the past two centuries”. Studies have looked, for example, at the long-term tendency for gerund-participial clauses to replace infinitival clauses (e.g. Fanego 2007; Vosberg 2009) and at synchronic variation between bare infinitival and *to*-infinitival complements (Callies 2013). A baseline of *to*-infinitives allows us to discount the possibility that the apparent decline in the *to*-infinitive perfect is simply a corollary of variation in the use of *to*-infinitival complementation in general.

To test this hypothesis we employed the search string “to [v?i*]” (*to* + infinitive verbs) to retrieve baseline frequencies, and then plotted the proportion p for all perfect cases, which we might write as $p(\textit{to}\text{-infinitive perfect} \mid \textit{to}\text{-infinitive})$. This obtains an almost identical result to that obtained with a per-million-word baseline. The observed decline is not due to an overall decline in *to*-infinitive frequency. On the other hand, the resulting plot is still apparently linear.

However, as discussed earlier, there is generally a temporal contrast between the *to*-infinitival perfect and a corresponding non-perfect infinitival structure, since the perfect expresses anteriority. A more reliable baseline for comparison purposes might be all past tense verbs (cf. Bowie, Wallis, and Aarts 2013a), which we can obtain using the simple search string “[v?d*]”. To this baseline we add the smaller set of present perfect constructions, because (as noted in Section 1 above) the meaning conveyed by the *to*-infinitival perfect can sometimes be rendered by

a finite construction containing a present perfect rather than a morphologically marked past tense.⁹ It can also be rendered in some instances by a past perfect, but those are already retrieved by our past tense search.

In this approach, we are considering the *to*-infinitival perfect as one of a group of verbal forms or constructions that can express anteriority. This is important because texts vary considerably in their temporal orientation (their concern with the past, present, or future). An infinitival perfect is likely to be used more frequently in texts with greater focus on past time, reflected in a greater use of past tense verbs. We therefore wish to test whether the apparent decline reflects a genuine change in preferences for the form, or reflects variation in the temporal orientation of texts sampled (which may not be controlled for). Our baseline represents a meaningful superordinate category for this purpose.¹⁰

The results of this comparison against past tense + present perfect are presented in Figure 3. This finds a smoother, more consistent decline against this baseline that approximates to a logistic regression line (popularly known as an “S-curve”, in this case with a close fit, $r^2 = 0.9603$).¹¹ Measured in this way the fall is huge: a decline of around 80 % from 1820 to 2000.

More generally, although the confidence intervals are small, joining the dots still obtains a “bumpy” line. The peak in 1880 is likely to be genuine (it appears in graphs with different baselines), but it is distinct from the overall logistic trend. This suggests that, for the *to*-infinitival perfect at least, the opportunity to use the form varies according to other uncontrolled factors, not just temporal orientation,

⁹ For the present perfect, we combined two searches using POS tags to allow for either a third-person singular or general present tense form of HAVE followed by a past participle: the strings “[vhz] [v?n*]” and “[vh0] [v?n*]”.

¹⁰ An alternative perspective, which has a similar outcome, focuses on a more specific potential alternation. This is applicable to examples that occur as complements to catenative verbs, which comprise a large majority of the total (see sections 3 and 4). We could consider the probability that a writer decided to insert a catenative verb and adjust the original tensed verb to become the complement of the catenative, i.e., to transform “subject + V-ed” → “subject + catenative verb + perfect infinitive *have* + V-en” (e.g. *he said* → *he seems to have said*). Similar transformations would also apply to “subject + present/past perfect”. In this kind of transformation, the meaning changes as a result of the insertion, but the decision to insert a catenative is not heavily constrained (except, naturally, as the sum of all meanings potentially expressible by catenative + perfect). The decision to use a particular catenative verb is one that will likely be sensitive to context, implying the need for additional analyses with different alternation patterns for each verb. The focus here is the overall trend, however, implying that all we need assume is that the total set of contexts that would permit insertion is stable over time.

¹¹ For more information on the method see <http://corplingstats.wordpress.com/2015/04/24/logistic-regression>.

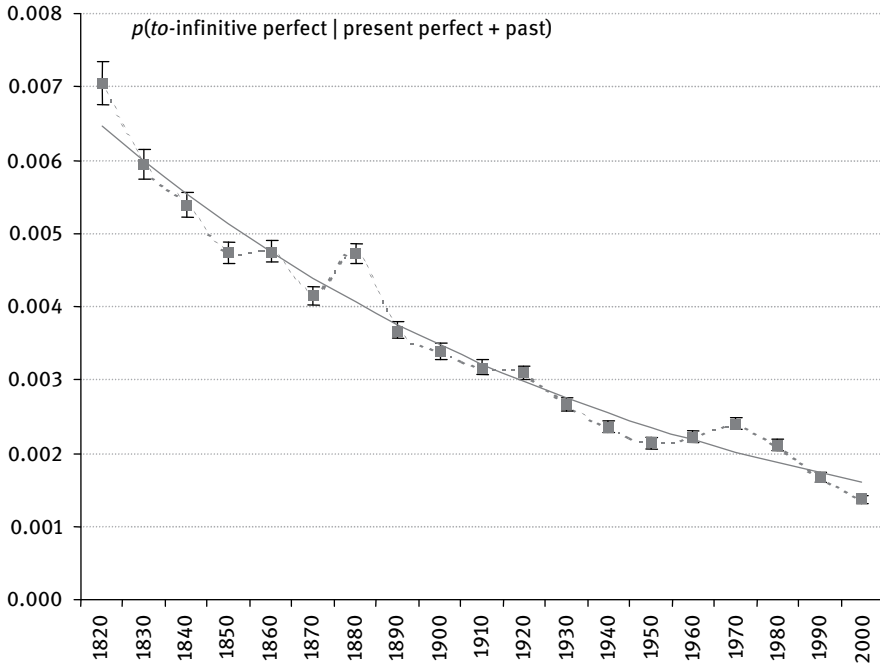


Figure 3: Relative frequency of the *to*-infinitival perfect against a baseline consisting of present perfect constructions plus past tense verbs, over the decades in COHA.

which might vary from text to text within the sample. It is necessary to dig deeper by examining the structural and semantic contexts in which the perfect is used.

We also compared the frequency of the *to*-infinitival perfect against the baseline of past tense verb plus present perfect constructions for each of the four genres in COHA. We found that all four genres show significant declines over the period, and we employed logistic regression to identify each underlying trend for the purposes of comparison.

The structure is consistently the least frequent in fiction, falling from around 0.4 % of cases in 1820 to a quarter of that rate by 2000, and declines less steeply in that genre than magazines and newspapers, which fall to as much as a seventh of their starting point. Although newspaper data is absent in the early decades, from 1860 onwards the remaining pattern is remarkably consistent with that for magazines. The category of “non-fiction books” sees the most gentle decline, from around 1% to 0.4% over the same period, becoming the genre containing the highest proportion of *to*-infinitival perfect cases from 1950 on.

3 Identifying contexts of occurrence

A further investigation of the decline in the *to*-infinitival perfect requires us to consider the contexts in which it occurs. *To*-infinitival clauses occur in a wide range of grammatical structures, as described in English reference grammars. These are structures within which the perfect form in question can potentially occur.¹² Huddleston and Pullum (2005: 212) provide a convenient list with examples, several of which are cited below, with the clauses in bold.

- (5) ***To turn back now** would be a mistake.*
- (6) *Her parents intend **to buy her a car**.*
- (7) *I go to the gym **to keep fit**.*
- (8) *We found a big box **to keep the CDs in**.*
- (9) *He was anxious **to make a good impression**.*

In (5) we have a subject; more common is an extraposed subject (*It would be a mistake **to turn back now***). The other functions illustrated are complement of a verb (6), clausal adjunct (7), modifier within a noun phrase (8), and complement of an adjective (9). Other functions listed by Huddleston and Pullum include extraposed object, complement of preposition, and complement of noun.

As can be seen from the above examples, it would be extremely hard to identify all such contexts in an unparsed corpus like COHA. Instead, we propose to determine in which contexts a perfect is most frequently found. To obtain a reasonable estimate of the frequency distribution for the perfect in COHA we use the parsed corpus ICE-GB, assuming that the differences between AmE and BrE are likely to be minor in this respect. Employing this relatively small parsed corpus allows us to do trial runs where we can obtain a complete set of results by automatic search before attempting to perform analogous searches on the unparsed COHA.

¹² This does not mean that they can occur in all instances of such structures. In particular, Huddleston and Pullum (2002: 161) note that “Most but not all catenative verbs allow their non-finite complement to be headed by the perfect auxiliary *have*.” For instance, they contrast *He seemed/*began to have offended them*.

3.1 ICE-GB

The *British Component of the International Corpus of English* (ICE-GB; Nelson, Wallis, and Aarts 2002) is a 1-million-word parsed corpus of BrE, synchronically sampled in the early 1990s and comprising both spoken and written material (60% and 40% respectively). It was grammatically analysed, using a phrase structure grammar based on Quirk et al. (1985), by a small team of linguist annotators at the UCL Survey of English Usage, and was first published in 1998. The parsing of all 84,000 “sentences” (i.e. punctuated sentences in writing and putative “sentences” in the speech data) involved a combination of automatic methods, specialist computer software, and patient manual correction, and took around 15 person years. Such an effort would clearly be impossible in the case of the 400-times larger COHA corpus.

For our present purposes, the main advantage of this corpus is not simply that the corpus has been parsed systematically according to a grammar scheme, but that human annotators have been able to address structures unanticipated by a formal grammar or parsing algorithm. They have also been able to disambiguate alternative possible analyses using pragmatic and semantic world knowledge. Human errors will undoubtedly exist, but are non-deterministic, unlike the errors of an automatic parser. From a statistical point of view, human annotation substantially improves overall precision by covering all cases, while errors do not introduce a systematic bias to results. Wallis (2014) discusses the classes of evidence that it is possible to derive from a parsed corpus: here we are interested in frequency evidence.

ICE-GB and its sister corpus, DCPSE, are supplied with the ICECUP research tool, which uses a powerful grammatical query system termed *Fuzzy Tree Fragments* (FTFs). These are visual grammatical query structures that the program matches against parsed trees and sentences in the corpus to retrieve results. The software visually exposes both the query and matched sentences, so that it is very easy to see how any given query matches, or fails to match, a particular tree in the corpus.

As a result, it is simple to experiment with numerous alternative queries and learn the grammar scheme as one works with the corpus. It is even possible to identify some portion of a parse tree and abstract an FTF from it. This “cyclical exploration” method allows novice and expert users to gain a high level of research competence and also permits grammarians committed to another scheme (e.g. that of Huddleston and Pullum 2002) to translate their query into the ICE (i.e. the Quirk et al. 1985) grammar framework to obtain data. Even when one framework does not have a one-to-one relationship with another, multiple queries can be performed. In summary, using ICE-GB as a test micro-corpus has

lost follows Quirk et al. by being labelled as “VP”, but excluding elements following the main verb such as objects and clausal adjuncts. The first search finds examples within a range of different structures, for example following a verb other than a semi-auxiliary (as in (11)) or following an adjective (as in (12)).

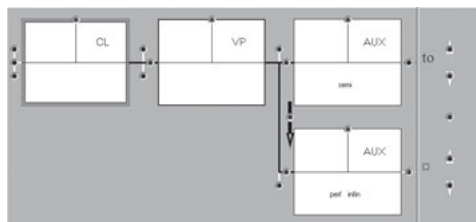


Figure 5: FTF for semi-modal *to*-infinitive structures.

- (11) *A scientist named Bridgeman **claimed to have shown** that there was a deviation from Ohm’s Law at very high current densities [...]*
(ICE-GB S2A-041)
- (12) *I am **sorry to have hounded** you over the past few days [...]*
(ICE-GB W1B-020)

In the present research we follow an analysis where auxiliary verbs (such as the modals and progressive BE) and “semi-auxiliaries” (like SEEM) are treated, together with verbs like CLAIM in (11), as *catenative* verbs, i.e. verbs taking a non-finite complement (see, e.g., Huddleston and Pullum 2002: 1209–1220 for a discussion of alternative analyses and arguments in favour of the catenative analysis). Therefore, for our purposes we combine the results of the two searches above. This obtains nearly 200 cases, which we examine and classify according to the function of the *to*-infinitival perfect clause. The combined results are presented in Table 1.¹⁴ The label “postmodifier” is used, as in the corpus analysis, to include both complements and modifiers within the phrase. The distinction is sometimes difficult to apply and is not necessary for present purposes.

The overwhelming majority of examples (84 %) function as complement of a verb. Another 11 % function as postmodifier within an adjective phrase or noun phrase, while the remaining 6 % fill several different clausal functions. This suggests that a collocational analysis of verbs occurring with the *to*-infinitival perfect

¹⁴ Four examples were excluded from the final counts as their status was dubious.

will be worthwhile: something that can be readily carried out in an unparsed corpus like COHA.

Table 1: Functions of *to*-infinitival perfect clauses in ICE-GB.

Function	N	%
complement of verb	161	84
postmodifier in AJP	14	7
postmodifier in NP	7	4
extraposed subject	6	3
adjunct	3	2
subject complement	1	1
Total	192	100

Catenative verbs taking *to*-infinitival complements can occur in different structural patterns. Further examination of the complement-of-verb examples from ICE-GB provided additional relevant information about the frequencies of different structural types.

We follow the analysis of Huddleston and Pullum (2002: 1177–1178), which divides catenative structures into “simple” and “complex” according to whether an intervening noun phrase occurs between the catenative verb and the *to*-infinitive in the active construction.¹⁵ This noun phrase can often become the subject of a passive in the matrix clause, and such structures are also included as “complex” catenatives although there is no intervening noun phrase in the actual surface structure. This gives us the three structural possibilities illustrated in (13)–(15).

- (13) *The Iraqis **claim to have shot** down fourteen aircraft.*
(ICE-GB S2B-008) (active, simple catenative)
- (14) *Does he **consider sampling to have been** an important innovation?*
(ICE-GB S2B-023) (active, complex catenative)
- (15) *Candidates are usually **expected to have reached** a standard well above the pass mark in their qualifying examinations.*
(ICE-GB W2D-007) (passive, complex catenative)

¹⁵ This is an area of the grammar where there are competing analyses of the grammatical status of the intervening noun phrases with different verbs. Discussion of these issues is beyond the scope of this paper. The division between “simple” and “complex” is, however, valuable for our present purposes regardless of the grammatical analysis chosen.

Example (13) is active and there is no intervening noun phrase, so this is a simple catenative structure. Example (14) is active and has an intervening noun phrase (*sampling*), so this is a complex catenative structure. Finally, (15) is passive and can be compared with an active counterpart with an intervening noun phrase (e.g. *We usually expect candidates to have reached...*), so is classified as a complex catenative structure. The frequencies of these structures are presented in Table 2.

Table 2: Catenative structures containing *to*-infinitival perfect clauses in ICE-GB.

Structure	N	%
active, simple catenative	126	78
passive, complex catenative	29	18
active, complex catenative	6	4
Total	161	100

The majority of examples (78%) involve simple catenative structures with the matrix clause in the active voice. Another 18% involve complex catenative structures with the matrix clause in the passive voice. This means that 96% have no intervening noun phrase in the surface string. This is useful information for planning data retrieval in an unparsed corpus like COHA, where it is difficult to allow for intervening noun phrases of widely varying lengths without also including irrelevant examples. For instance, consider the intervening noun phrase (indicated in square brackets) in the following example:

- (16) *He **estimated** [the slope at the time in nineteen eighty-four] **to've been** something like one in four.*
(ICE-GB S2A-067)

In our parsed corpus, searches can allow for any intervening noun phrase, even one longer than the eight-word string above, but this is not possible in an unparsed corpus.

Finally, we also identified the verbs which occurred in the 161 examples of verbs taking *to*-infinitival complements with the perfect. There were 29 different verb lexemes, but the two most frequent accounted for over half of the data: SEEM (34%) and APPEAR (22%). The next most frequent verbs, OUGHT and CLAIM, accounted for only 4% each. This suggests that searches for the most frequent verb collocates of our target structure in COHA will enable us to retrieve large proportions of the relevant structures.

4 Collocations and semantic categories in COHA

Our ICE-GB results lead us to expect that the great majority of *to*-infinitival clauses containing a perfect auxiliary in COHA will occur as the complement of a verb, i.e. in catenative constructions (the proportion in ICE-GB being over 80 %). We further expect that only a small minority of these catenative constructions will include an intervening noun phrase between the catenative verb and its infinitival complement (the proportion in ICE-GB being only 4 %).

We cannot specify the proportions in COHA as precisely as for ICE-GB, but we can give reasonable estimates by modifying our original search string for the *to*-infinitival perfect. If we add an immediately preceding verb to the string (using the tag “[v*]” to find any verb), the results comprise 67 % of the total found by the original string. That search excludes not only those catenative constructions with intervening noun phrases, but also those with other intervening material between the catenative verb and *to* (e.g. *seems to me to have, claimed not to have*). We estimate that genuine examples of these kinds comprise no more than another 10 % of the total found by the original string with no preceding verb.¹⁶ Thus the overall proportion of *to*-infinitival perfects occurring in catenative constructions (around 77 %) seems to be rather lower than in ICE-GB, but of a similar order of magnitude.¹⁷

As we saw with ICE-GB, we have good reason to believe that the lexical semantics of the preceding catenative verb paradigmatically frames possible alternation patterns. Our next step therefore involved identifying these collocates.

We retrieved the left-hand verb collocates of the *to*-infinitival perfect string using the corpus interface “List” function – search string: “to have [v?n*]”; collocates: “[v*]” (i.e. any verb), left: 2, right: 0; group by lemmas; increase hits frequency above the 100 default until a full list is displayed (see Appendix B for a screenshot).¹⁸ This obtained a list of 402 verb lemmas (or lexemes) in order of declining frequency, with a total number of 44,321 tokens. This, as one would expect, is almost identical to the number of tokens retrieved by the string search

16 This estimate is based on an examination of data retrieved by search strings that allow for intervening material of various kinds in this position, with account taken of irrelevant examples found by such searches. Note that these comments on proportions apply to COHA taken as a whole. Strictly, we cannot assume that the frequencies of the patterns remain completely stable over this 190-year period. For instance, while the overall frequency in COHA of our *to*-infinitival string immediately preceded by a verb is 67 %, the proportions per decade range from 59 % to 71 %, with a tendency to gradually increase over time.

17 Recall that ICE-GB represents 1990s BrE, with 60 % being spoken data.

18 That is, it was necessary to experiment with increasing the default hit frequency of 100 until no further examples were found by further increasing the limit (in this case, to 500, say).

“[v*] to have [v?n*]” mentioned above; the data again exclude those examples with intervening material between the verb and *to*. Examples include both active, simple catenative structures and passive, complex catenative structures.

The results of our search present a lexical profile of the *to*-infinitival perfect in the catenative construction. Table 3 lists the top 30 lemmas in rank order with individual frequencies, cumulative frequencies, and cumulative percentages of the overall data. These 30 items account for 95 % of retrieved tokens – the “top 95 percentile” of the data. We reasoned that although the remaining “long tail” of 5 % of tokens may contribute to the overall pattern of change, precisely because

Table 3: Catenative verb immediate collocates of the *to*-infinitival perfect in COHA: top 95 percentile.

Rank	Lemma	Category	Frequency	Cumulative	%
1	SEEM	seeming	17,174	17,174	39
2	APPEAR	seeming	5,060	22,234	50
3	SAY	cognition and saying	4,964	27,198	61
4	OUGHT	modality	3,998	31,196	70
5	BE	modality	1,791	32,987	74
6	REPORT	cognition and saying	1,271	34,258	77
7	CLAIM	cognition and saying	970	35,228	79
8	BELIEVE	cognition and saying	957	36,185	82
9	KNOW	cognition and saying	953	37,138	84
10	LIKE	–	594	37,732	85
11	THINK	cognition and saying	429	38,161	86
12	UNDERSTAND	cognition and saying	420	38,581	87
13	ALLEGE	cognition and saying	399	38,980	88
14	REMEMBER	–	385	39,365	89
15	FIND	cognition and saying	334	39,699	90
16	HAVE	modality	264	39,963	90
17	PROVE	–	256	40,219	91
18	HAPPEN	–	244	40,463	91
19	PROFESS	cognition and saying	165	40,628	92
20	PRETEND	–	156	40,784	92
21	SHOW	cognition and saying	150	40,934	92
22	CONSIDER	cognition and saying	140	41,074	93
23	REPUTE	cognition and saying	121	41,195	93
24	MEAN	prospective meaning	120	41,315	93
25	WISH	prospective meaning	119	41,434	93
26	ESTIMATE	cognition and saying	115	41,549	94
27	PRESUME	cognition and saying	112	41,661	94
28	EXPECT	prospective meaning	106	41,767	94
29	HOPE	prospective meaning	102	41,869	94
30	SEE	cognition and saying	98	41,967	95

they are low in frequency, they are unable to explain the *scale* of the change seen in Figure 3. If they declined to zero occurrences, they could only explain at most a 5% decline.

We assigned most of the top 30 collocates to groups based on their meaning and grammatical properties, as summarised below. Five verbs were not assigned to any grouping, either because of grammatical or semantic differences, and/or because they showed two distinctly different uses in the data and so did not fit any one group. The rationale behind the groupings will be explained below. The distribution of the lemmas is plotted in Figure 6, and the distribution of the groups in Figure 7.

By far the most frequent collocate is *SEEM*, accounting for 39% of the data. The next most frequent is *APPEAR*, with 11%. Thus, these two semantically similar items account for half the data.¹⁹ Similar to *APPEAR* in frequency is *SAY* (11%), followed by *OUGHT* (9%). After that there is another steep drop to *BE* (4%) and thereafter an exponential decline. This distribution, visualised in Figure 6, can be thought of as an example of Zipf's law applied to a single grammatical "slot".

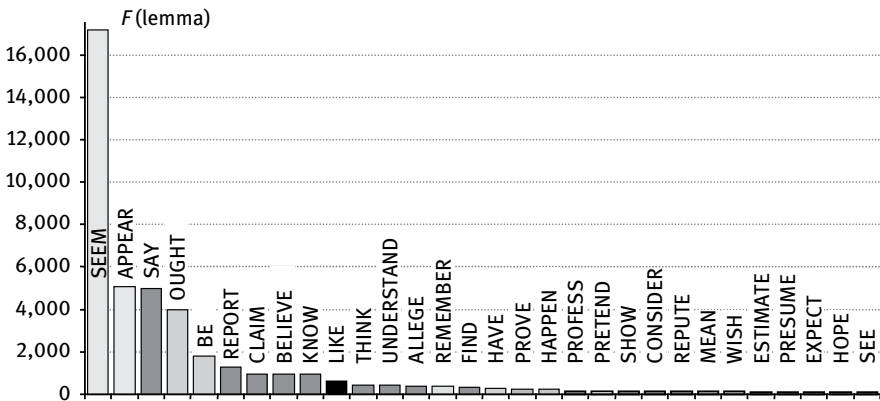


Figure 6: Frequency distribution of the top 95 percentile of catenative verb lemmas (as in *SEEM to have*, etc.) in COHA (all data).

It is notable that many of our top 30 verbs are broadly concerned with the areas of evidentiality and modality: i.e. with indicating the source or reliability of the evidence for a proposition, or expressing the speaker's commitment to the truth

¹⁹ These two items were also the most frequent in ICE-GB, where the share for *SEEM* was 34%, while that for *APPEAR* was 22%, twice as high as in COHA.

of a proposition or attitude to the desirability of a situation being actualised. They are concerned with what is believed, known, said or reported to be the case, what appears to be the case or is shown to be the case, with what should be done, and so on. A related semantic-grammatical property shared by a striking number of the verbs is that of occurring with a raised subject (either in the active, or in the passive where it is a “derived” raised subject).

A raised subject differs from an ordinary subject in that it does not have a semantic role in relation to the matrix verb, only having one in relation to the verb in the subordinate clause, and this semantic distinction is reflected in various aspects of grammatical behaviour (see, e.g., Huddleston and Pullum 2002: 1194–1198, 1200–1203).

For example, the “seeming” group verbs take a raised subject in the active, simple catenative structure (described in Section 3.2). In *He appears to know them*, it is not “he” who appears, but rather “he to know them” that appears to be the case (compare *It appears that he knows them*, with *it* as “dummy” subject). It is as if *he* has been “raised” into subject position out of the lower clause where *he* semantically belongs. In contrast, a verb like *offer* takes an ordinary subject. In *He offered to help us*, the subject *he* has a semantic role in relation to *offered* as well as to *help*.

The verbs in the “cognition and saying” group take a raised object, which differs from an ordinary object in an exactly parallel way. Thus, in *She believed him to be dead*, it is not “him” that she believes, but rather “him to be dead” (compare *She believed that he was dead*). It is as if *him* has been “raised” to object position out of the lower clause where *him* semantically belongs. By contrast, in *She persuaded him to leave*, *him* has a semantic role in relation to *persuaded* as well as to *leave*. In the passive counterpart of our raised-object example, *He was believed to be dead*, we can consider *he* to be a “derived” raised subject. Such passives are frequent in our data.

The raised subject property seems to be semantically connected to the “modalising” or “evidentialising” contribution that many of the collocate verbs make in relation to the situation described in the infinitival complement clause. We can see this kind of contribution by comparing a simple construction like *He was dead* with *He seemed to be dead* or *He was believed to be dead*. The catenative verbs relate semantically to the entire situation of the complement clause, and do not assign any further semantic roles to explicitly expressed participants; there is an implicit “believer” role in the example with *believed*, but its importance is downplayed by its being left implicit. By contrast, if we passivise our ordinary-object example to get *He was persuaded to leave*, we see that the verb *persuaded* does assign a semantic role to the subject *he* (the same role as when it is the object in the active counterpart).

In categorising and describing our verbs, we have taken account of properties used or noted in the classification of catenative verbs provided in Huddleston and Pullum (2002: 1225–1245), including:

- occurrence in the simple and complex catenative constructions (see Section 3.2 above): simple only, both simple and complex, or complex only,²⁰
- raised versus ordinary subject or object,
- possibilities for passivisation,
- alternation with structures involving finite subordinate clauses.

Huddleston and Pullum also cover the formal type of non-finite complement taken (*to*-infinitival, bare infinitival, gerund-participial, past-participial, or more than one of these). This dimension will be largely ignored here, as our focus is on *to*-infinitivals (although the possibility of alternation with a gerund-participial clause is relevant for some verbs).

A further relevant property concerns the temporal interpretation that catenative verbs allow for their complement in the absence of the perfect, as discussed elsewhere in Huddleston and Pullum (2002: 160–162). For instance, they contrast the interpretations of *I remembered/enjoyed/intended going to Sydney at Christmas*. The time of going to Sydney at Christmas is anterior to that of remembering, simultaneous to that of enjoying, and posterior to that of intending (where the going may not have been realised). With some verbs, the choice of temporal interpretation depends on the aspectuality of the complement clause. For instance, compare *She may be in her office* (imperfective: simultaneous) and *She may resign* (perfective: posterior). The temporal interpretation of the non-perfect is relevant to the effect of applying the perfect.

It should be noted that some verbs have two or more different uses or meanings which are assigned to different categories in Huddleston and Pullum's classification. This is the case for some of our collocate verbs, as we note below. In some instances, such divided usage means we have not assigned a verb to any category; in others, a verb has been assigned to a category based on the use involved in the great majority of examples.

We used the above criteria, along with other semantic properties, to formulate principal semantic-grammatical classes (here termed “verb groups”) for each verb lemma, taking the most frequent verbs first. Although we are not principally

20 For our purposes, we are considering occurrence in the simple and/or complex catenative constructions with a *to*-infinitival complement only. Huddleston and Pullum's classification considers other formal types of non-finite complement also (e.g. gerund-participial). A verb can occur in both simple and complex constructions, but with different formal types of complement (e.g. gerund-participial and *to*-infinitival).

interested in the possibility of alternation between lemmas (e.g. choosing SEEM over APPEAR and vice versa), by grouping categories in this way we attempt to isolate contexts of variation between the *to*-infinitival perfect and other structures, thereby allowing us to pose the question: *where are the perfects going?*

These categories, like the lemmas, are similarly exponentially distributed. Half the data is explained by the first, “seeming” group, and most of the remainder falls under the next three groups. Figure 7 summarises the top of this distribution.

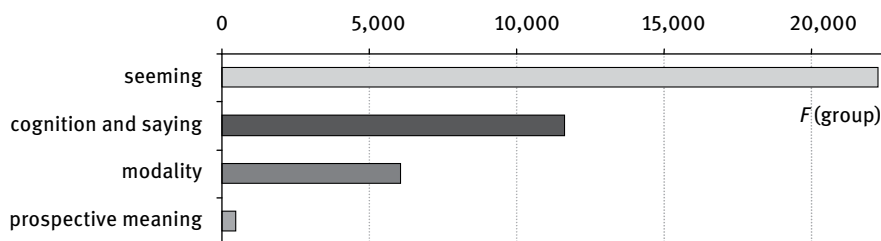


Figure 7: Frequency distribution of the top four verb semantic groups, COHA (95 percentile, all data).

The proportion of cases in each of these semantic groups is not constant over time, however, as Figure 8 reveals. There may be a number of potential reasons why the “seeming” category, while consistently dominant, falls to a minimum in the 1930s before rising again, whereas the three “modality” verbs peak in the 1870s. In each group there are two types of potential explanation: that the contextual *opportunity* for writers to use this type of expression varies, or that the decision to express “seeming”, “modality” etc. *with the perfect* is varying. This change in share is substantial. Both the “modality” and the “cognition and saying” groups vary by as much as 100 % (10–20 % and 20–40 % respectively).

However, we also know that this variation within the share of perfects is found within a declining form (see Figure 3). Figure 9 allows us to visualise the decline of each group relative to the decline of the perfect *to*-infinitive, expressed here by the 95 percentile “Total” line. From this perspective, we can see different verb groups decline over different periods and at different rates: first “seeming” (1820–1930: ~60 % fall), then the “modality” group (1880–1950: 75 %), and finally “cognition and saying” (1970–2000: 60 %, alongside a further 40 % decline in “seeming”). The probabilities of each group (or individual verb) sum to the total.

In the following subsections we look at each of these groupings in turn, and then consider the remaining five uncategorised verbs in the 95 percentile.

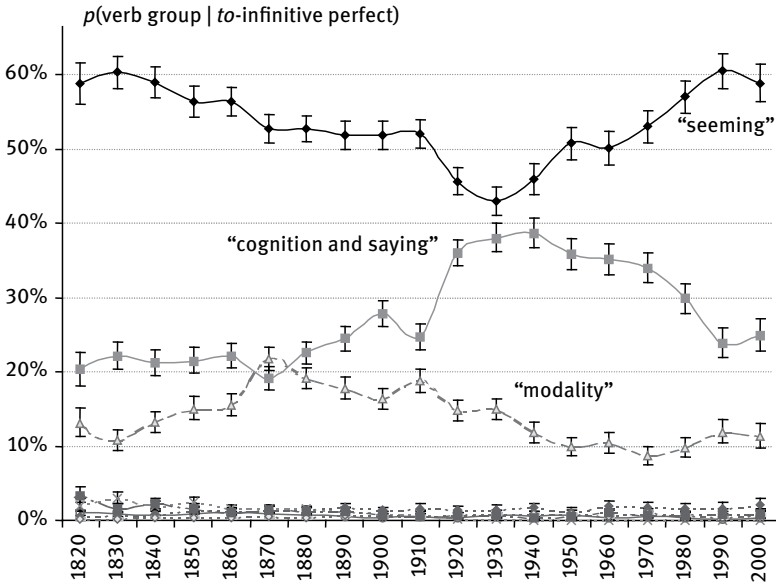


Figure 8: The changing meaning of the *to*-infinitive perfect: variation in the semantic distribution of the catenative verb group over time (as proportion of the top 95 percentile).

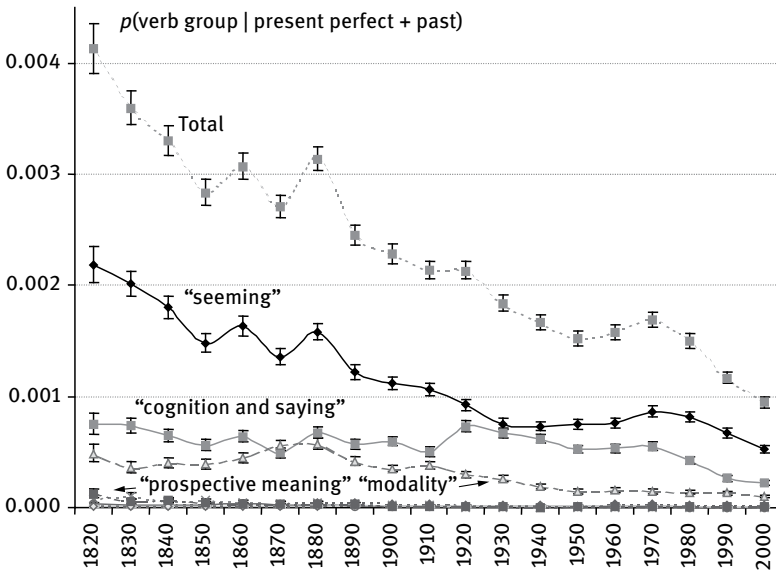


Figure 9: Perfect meaning variation (95 percentile total, catenative verb groups and remaining lemmas) plotted against the present perfect + past baseline.

4.1 The “seeming” group

SEEM and APPEAR are both semantically and grammatically similar. Here are a couple of examples from the corpus, with the catenative verbs in bold:

(17) *His old age **seems** to have been as quiet and happy as his youth was restless and miserable.*
(1821:MAG NorthAmRev)

(18) *By then, the addiction **appeared** to have ended Howe’s ability to earn major league money again [...]*
(1991:MAG SportsIII)

Both verbs occur in the simple catenative construction only and take a raised subject. In the absence of a perfect infinitive to express anteriority, the time of the complement situation is interpreted as being simultaneous with that of the matrix situation (e.g. *He seems/appears to like her*). Both verbs can also take a finite subordinate clause in the impersonal construction.²¹

Compare the above examples with the following:

(17’) *[...] it **seems** that his old age was as quiet and happy as his youth was restless and miserable.*

(18’) *[...] it **appeared** that the addiction had ended Howe’s ability to earn major league money again.*

HAPPEN is the only other verb in the top 30 list which shares all the above properties, but as it is different in meaning and far less frequent than SEEM and APPEAR we have not included it in the same group.

We will first plot the proportion of both verbs out of the present perfect + past baseline over the time period. Figure 10 shows that both falls are reasonably consistent with a logistic decline ($r^2 = 0.8952$ and 0.7966 respectively).

²¹ This construction is sometimes treated as an instance of extraposition, but differs in that the subordinate clause cannot fill the actual subject position: **That Dunlap was governed chiefly by caprice seems*. See Huddleston and Pullum (2002: 960–962).

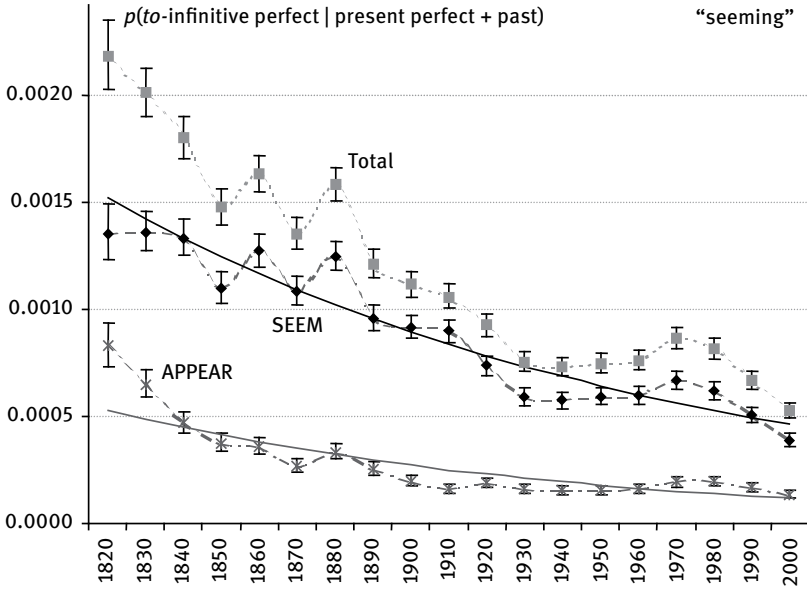


Figure 10: “Seeming” perfect *to*-infinitive over time, as a proportion of the present perfect + past baseline, COHA.

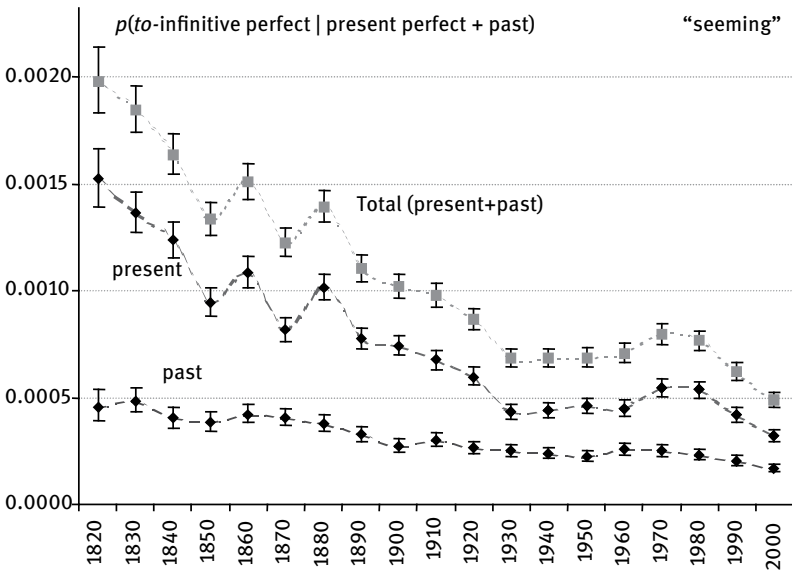


Figure 11: “Seeming” verbs, present tense and past tense, plotted against the present perfect + past baseline, COHA.

Although these appear to be approximately parallel, it turns out that the ratio between SEEM and APPEAR in the *to*-infinitive perfect is not constant over time. SEEM increases its share of the “seeming” verbs from 62 % to 85 % between 1820 and 1910, before stabilising and gradually declining from 1950 onwards.

Present tense forms of SEEM comprise 64 % of the data, against 28 % for the past tense and just 8 % for other, non-finite forms. So another way of subdividing the data would be to examine present and past tense forms.

Figure 11 compares the *to*-infinitival perfect decline following present and past tense forms with the total of the two. This shows a steeper decline with present tense contexts than past tense (over the period 1820–2000: around 80 % for present, 60 % for past). As a result, the proportion of past forms climbs from around 25 % of all “seeming” perfects in 1820 to about 35 % by 2000.²²

How might we study the alternation of individual verbs between perfect and non-perfect patterns? The non-perfect is not in general able to replace the perfect in the infinitival complements of the seeming verbs without a change of meaning (involving loss of anteriority). For instance, *His old age seems to be quiet and happy* would imply he is still alive and enjoying his old age, unlike (17) above (with *seems to have been*). Nor, to take an example with a non-stative lower verb, could we replace *The train seems to have stopped* with *The train seems to stop*.

We therefore need to explore other possible alternates. One possibility is that finite complement clauses, as in (17') and (18') above, are being used instead of infinitival complement clauses. It is not possible in an unparsed corpus to retrieve full and precise data for finite complement clauses containing a past tense verb or present perfect construction, but we were able to use multiple search strings to obtain an indicative sample of these data.²³

For both SEEM and APPEAR considered separately, we evaluated the probability of selecting the *to*-infinitival perfect rather than the finite construction, i.e. the probability of selecting examples like (17) rather than (17') and (18) rather than (18'). We found that SEEM and APPEAR behave slightly differently: the probability of choosing the *to*-infinitival perfect with SEEM declines from 85–90 % to around 75 % of cases, with the main fall from 1950 to 2000; with APPEAR, on the other hand, the probability is relatively stable. The picture is clearly not one of simple replacement: this alternation cannot explain the decline in the *to*-infinitival perfect with the “seeming” verbs.

²² All three lines fit logistic curves with approximately 90 % accuracy, but this does not mean that past and present have a constant ratio.

²³ Our searches allowed for examples with various subordinating conjunctions (*that, like, as if/though*) or none, and were restricted to those with pronoun subjects, which gave more reliable results.

If they are not subject to simple replacement by finite complements, what is happening to these *to*-infinitive perfect “seeming” cases? Authors may be choosing to express this type of meaning less frequently or they may be finding other ways to express similar meanings. Possible alternative forms include adverbs such as *seemingly* and *apparently* (cf. *Apparently, the addiction had ended Howe’s ability...*), or parenthetical finite clauses with SEEM (as in *He had forgotten, it seemed*), to mention just a few. As acknowledged earlier, there is no clear limit to the set of possibilities, and investigation of them all is beyond the scope of the current paper.

4.2 The “cognition and saying” group

We categorised 16 verbs as belonging to what we termed a “cognition and saying” group, borrowing a term used informally by Huddleston and Pullum (2002: 1234). This group accounts for the second largest share of the data after the two “seeming” verbs. Most of these verbs occur in the complex catenative construction only, take a raised object, and tend to occur frequently in the passive. In fact, the passive is obligatory with some of these verbs, including SAY, the highest-ranking of this group in our top 30 list. Here are a few examples from our data:

- (19) *She was **said** to have become insane on receiving the news of her son’s shipwreck [...]*
(1867:FIC AmbroseFecitThe)
- (20) *He is **reported** to have enlisted at Marseilles [...]*
(1844:FIC EdmondDantès)
- (21) *More than 2 million North Koreans are **believed** to have died of famine from 1994 to 1997.*
(2003:NEWS USAToday)
- (22) *If a man is **known** to have killed another, the killer likewise is put to death.*
(1941:MAG NatGeog)

Recall that our COHA search string could only find immediate verb collocates of the *to*-infinitival perfect, which does not allow for an intervening noun phrase as typically found in an active complex catenative construction. As a result, passive constructions dominate the data for this group of verbs to an even greater extent than otherwise, so that constructions with “derived” raised subjects are domi-

nant. Like the “seeming” verbs, these verbs can also occur in constructions with a finite subordinate clause complement and dummy *it* as subject (extraposition constructions in this case). Compare (19) and (20) above with (19’) and (20’).

(19’) *It was said that she {became/had become} insane on hearing the news of her son’s shipwreck.*

(20’) *It is reported that he (has) enlisted at Marseilles.*

The “cognition and saying” verbs are also similar to the “seeming” verbs in that the time of the catenative complement clause is, in the absence of the perfect, interpreted as simultaneous with that of the matrix clause (e.g. *She was said to be living in Paris; He is known to be in the building*).

We do find some active, complex constructions in our data, in instances where a potentially intervening noun phrase occurs in an “extracted” position, due for instance to relativisation (23) or to preposing to pre-subject position (24).²⁴

(23) *The man whom you **believe** to have stolen a million is my own brother.*
(1869:FIC TriflesChristmas)

(24) *This person he **knew** to have been an active confederate [...]*
(1835:FIC HorseShoeRobinson)

While most of the verbs in this group occur with a *to*-infinitival complement in the complex catenative construction only, there are two which can occur in either simple or complex constructions: CLAIM and PROFESS. The overwhelming majority of our data for these verbs involves the simple construction, as in (25) and (26); example (27) shows one of a small number of passives found for CLAIM (none were found in the data for PROFESS).

(25) *He **claims** to have been at work in the Park stables at the time of the assault.*
(1885:NEWS NYT-Reg)

(26) *[...] Dickens **professed** to have been greatly inspired and influenced by Fielding and Smollett [...]*
(1925:MAG Harpers)

²⁴ The use of terms like “extracted” and “preposing” is not intended to imply a literal derivation in the speaker’s mind which starts from an underlying structure with the noun phrase in question in the intervening position.

- (27) *Take one of the principal points of argument, the evil which is **claimed** to have resulted from the attempt to study so many of the practical sciences.*
(1851:MAG NewEngYaleRev)

The simple construction with these verbs involves an ordinary rather than a raised subject, so they differ from the other verbs in this group where most of the examples involve passive constructions with (derived) raised subjects. However, they bear semantic similarities to other members of the group, and share with them the ability to occur with a finite subordinate clause, and the temporal interpretation (in the absence of the perfect) of simultaneity of the infinitival subordinate clause and the matrix clause. We have therefore included them in this group.

Note that the verb SEE is included in this group because, in the kind of usage in our data, it often has the sense of mental inference rather than sensory perception, and in this usage has similar grammatical properties to the rest of the group, as pointed out by Huddleston and Pullum (2002: 1236–1237). This usage is illustrated in (28).

- (28) *Projected into the past, the Earth's rotation is **seen** to have been a very little faster than its present rate.*
(1951:MAG Harpers)

A number of other verbs in the group seem to go beyond the central semantic area of “cognition and saying” to refer to processes of examining or presenting evidence for something: FIND, SHOW, and ESTIMATE. This is a semantically linked area, and they share the other properties of the group. Examples are given in (29) to (31).

- (29) *Roger was asked to leave college because he was **found** to have been black-mailing the wife of one of the business men in town.*
(1932:FIC MardiGrasMurders)
- (30) *Many were themselves accused of witchcraft; and noble ladies were **shown** to have dabbled in mystic arts [...]*
(1869:NF MemoirsExtraordinary)
- (31) *The value of the ore shipped from Creede in February is **estimated** to have been \$850,000.*
(1893:NEWS NYT-Ed)

The first task in examining variation in this group is to plot it over time against the general present perfect + past baseline, as in Figure 12. Against this baseline, the total for the group falls to about 30 % of its 1920 value (it declines by 70 %) from 1920 to 2000.²⁵

Over the same period, the predominant form, *SAY*, which initially accounts for half the forms in the group, declines to a quarter-share by 2000. The result is an increase in diversity of other forms, with first *REPORT* rising (1910–1940: +150 %) before falling, and then *CLAIM* rising (1950–1980: +75 %), to the point that *CLAIM* and *SAY* are not significantly different in frequency by 2000.²⁶

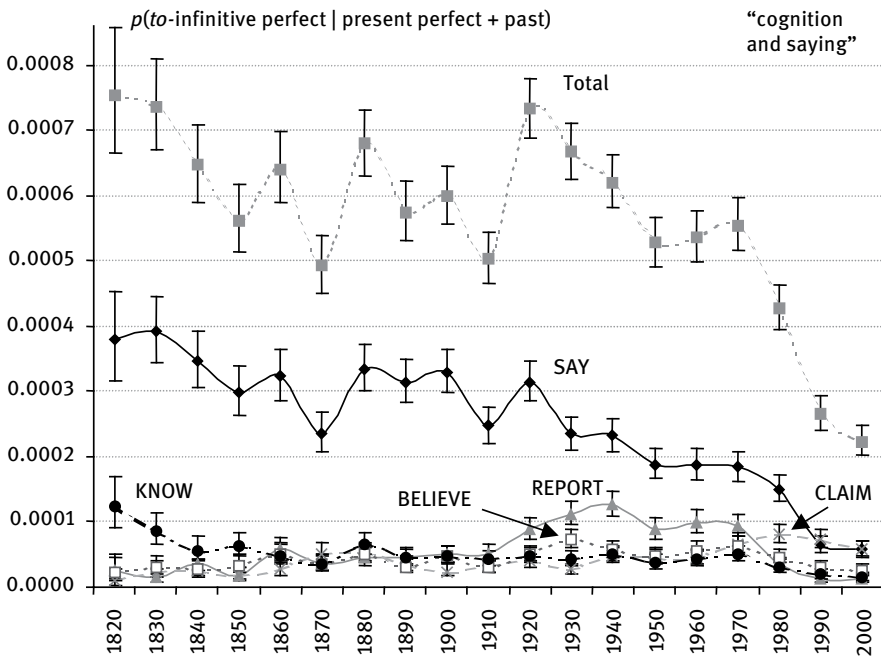


Figure 12: “Cognition and saying” perfect *to*-infinitive over time, against the general present perfect + past baseline, in COHA.

²⁵ Before 1920, the total line fluctuates, possibly due to text-sampling variation, but is most meaningfully interpreted as being reasonably steady. The same cannot be said for the sharp decline in the twentieth century.

²⁶ Note that with multiple competing forms, we should not be surprised that these do not all follow a simple logistic “S-curve” rise or fall. See <http://corplingstats.wordpress.com/2015/05/27/logistic-multinomial>.

Given that we have a complex pattern of change here, it may be worth considering whether in fact verbs are more or less likely to be used in a perfect construction than in a general *to*-infinitive. That is, suppose we compare *SAY* in the perfect *to*-infinitive (e.g. *said to have been*) versus the same verb with a non-perfect *to*-infinitive (e.g. *said to be*). Are some verbs more likely to be used in the first construction than other verbs?

If we examine the distribution of lemmas synchronically (i.e. for all periods of COHA combined), we obtain the distribution in Figure 13. For the purposes of comparison, we have maintained the same ordering as in Figure 6, i.e. *SAY* has more *to*-infinitival perfect cases than the others, etc. We can see that about 50% of *to*-infinitival cases of *ALLEGE* and *REPORT* are found in perfect constructions, which we might think of as a simple measure of their “attraction” to the perfect.

If we reorder this synchronic sequence according to attraction probability, and then take the difference in rank between the two sequences, we find that *ALLEGE*, *REPUTE* and *ESTIMATE* increase in rank most substantially, and *THINK* and *FIND* fall. These verbs are relatively infrequent and so do not affect the overall pattern of diachronic change plotted in Figure 12. Nonetheless, they indicate that there is a lot of variation within this one group, and that some verbs may be co-occurring with the perfect as an “idiom chunk”.

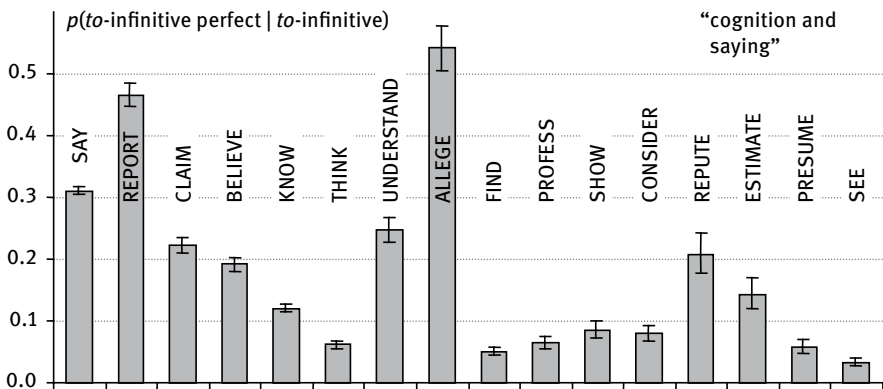


Figure 13: Plotting “attraction” to the perfect. Synchronic distribution of the probability of each “cognition and saying” verb being used in a perfect *to*-infinitival construction, given it is used in a *to*-infinitival construction. Comparison of column heights against the 95% Wilson score intervals indicates that this variation is wide and substantial.

As *SAY* is the dominant form and falls so sharply, we tested this verb for possible alternation of the *to*-infinitival perfect with a finite complement clause contain-

ing a past tense or present perfect (cf. (19') above). An indicative sample of such clauses was obtained by using several search strings. The results of the comparison do not show a consistent downward trend. Therefore, the alternation cannot explain the decline observed for SAY + *to*-infinitival perfect. Rather, this decline seems to be the result of the BE *said to*-infinitival perfect halving its share of all *to*-infinitival perfect forms over the period 1940–2000.

4.3 The “modality” group

The “modality” group accounts for the third largest portion of data. It includes the verbs OUGHT, BE, and HAVE, which rank at 4, 5, and 16 on the list respectively (see (32) to (34)).

(32) “[...] You **ought** to have let him do as he wished with his own wigwam. You were unjust”.
(1857:FIC RolloInWoods)

(33) He **was** to have come to me for a music lesson this afternoon, but as usual he sent an excuse.
(from dialogue in a play; 1879:FIC LastWord)

(34) To give examples I should **have** to have taken notes [...]
(1896:FIC Embarrassments)

These verbs, like the “seeming” verbs, occur only in the simple catenative construction, and have raised subjects. However, they differ from that group, and from the “cognition and saying” group, in that they do not take finite clausal complements. They also differ from those groups in terms of possibilities for temporal interpretation of the infinitival complement in the absence of the perfect. There are complex patterns involved, but it is noteworthy that the interpretation often involves posteriority rather than simultaneity (e.g. *We ought/are/have to visit her*).

Each of the three “modality” verbs has changed in frequency over the two centuries captured by COHA. Figure 14 depicts the proportion of the past + present perfect baseline for these three verbs. *Was/were* is plotted here, as past tense forms account for the great majority of the data for BE; see further below.

The variation in *ought* (the predominant form) explains the verb group pattern until the 1950s, with the other two forms being nearly constant. In the second half of the twentieth century, HAVE increases from nearly zero occur-

rences to converge with the falling *ought*. Meanwhile *was/were* declines by 70% from 1920 to 1990.

As these three verbs are rather disparate in meaning, we consider each in turn.

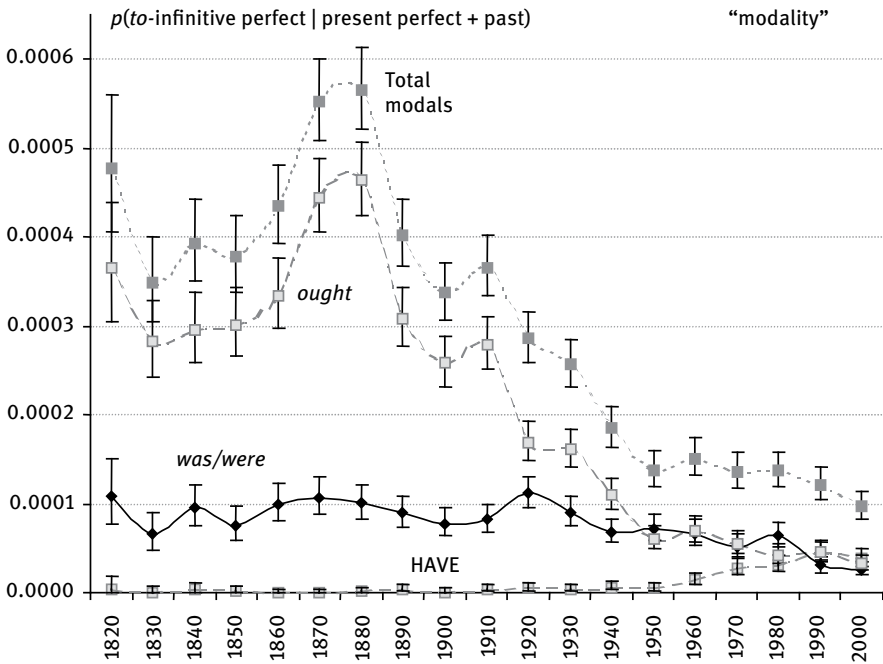


Figure 14: The rise and fall of the *ought to-infinitive perfect*, and other “modality” verbs in COHA.

4.3.1 OUGHT

Ought is considered to be primarily deontic in meaning, though it can also have epistemic readings (e.g. Collins 2009: 53–55); many of our examples, like (32) above, are clearly deontic in meaning, but there are some which are not, as in (35).

- (35) “[...] We really **ought** to have smashed to pieces. It’s strange that the ship withstood it”.
(1989:FIC Eden)

This conveys that one would have reasonably expected the ship and its passengers to smash to pieces, given the conditions: an epistemic, or perhaps dynamic (circumstantial), reading. *Ought* can convey an implication of non-actualisation, as seen in our examples with the perfect: *you* in (32) did not let him do as he wished; *we* in (35) did not smash to pieces.

An obvious semantic competitor for *ought* is *should*, which, as a modal verb, takes a bare infinitival complement. These items are generally viewed as being semantically very similar, and often interchangeable (e.g. Collins 2009: 52–53). We examine this alternation in Figure 15.

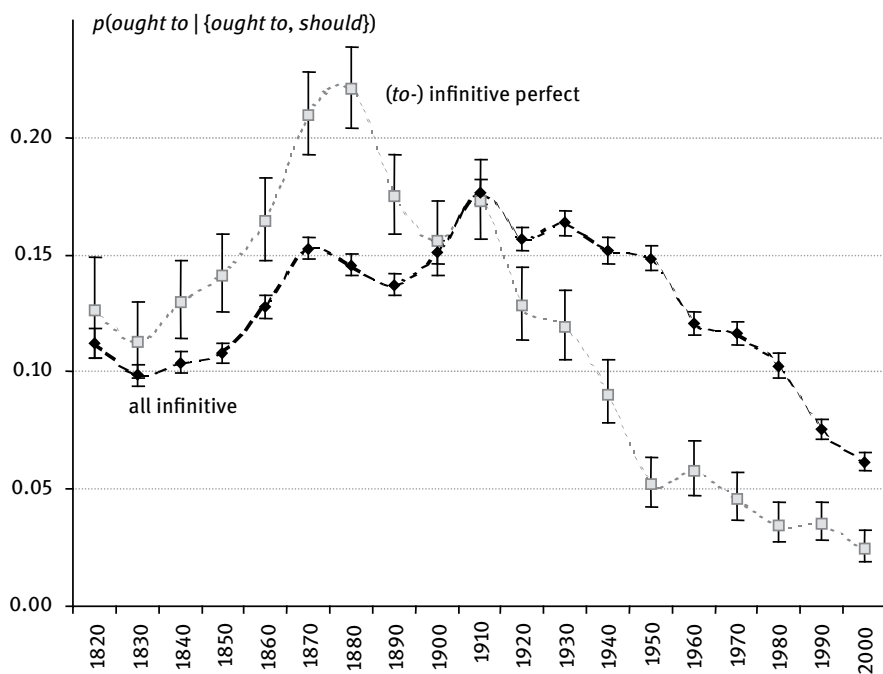


Figure 15: The rise and fall of *ought to* against its modal alternate *should* in COHA.

The dotted line represents the alternation within the perfect (*ought to*-infinitive perfect vs. *should* bare infinitive perfect). The dashed line represents the corresponding infinitive alternation. The baseline in each case is the set of both alternates, *ought to* and *should*.

The perfect line correlates with the *ought* line in Figure 14, suggesting that the *ought to*-infinitive is indeed seeing a pattern of replacement. In this pattern, *ought to* increases relative to *should* over the period 1830–1880 before declining in the

twentieth century. This peak presages a parallel pattern of modal replacement in the infinitive form overall, which outnumbers the perfect by around 10 : 1, shown as a darker dashed line in Figure 15.

4.3.2 BE

BE²⁷ can be used to express deontic necessity (e.g. *You are to come home at once*), but the examples typical of our data, like (33) above, express meanings involving schedules or plans – meanings that are perhaps “more temporal than modal”, in the words of Collins (2009: 85). As Collins notes, the perfect can be used to imply non-actualisation: this certainly characterises many of our examples, including (33) above, (36) and (37).

(36) *“I see not here your son, young Henry Winthrop,” said the Lady Arabella. “I thought he **was** to have been of our company”.*
(1846:FIC MyrtisWithOther)

(37) [...] *she had been left by a man the very day she **was** to have been married.*
(1909:FIC ThreeLives)

Often, the non-actualisation is obvious from context. Alternation with a non-perfect seems possible with examples like these, at least in present-day English, e.g. *I thought he was to be of our company; [...] the very day she was to be married.*

To test this possible area of alternation, we examined the proportion over time of *to*-infinitival clauses that contained a perfect when preceded by *was/were* (since BE occurs in the past tense in the great majority of our examples with the *to*-infinitival perfect: about 1,600 out of 1,791 examples). This does not show a clear diachronic trend, so does not support the hypothesis that, over time, speakers have increasingly chosen to use a non-perfect in such examples. BE followed by the *to*-infinitive shows a steady decline per million words in COHA during the twentieth century, which amounts to a decline in a potential context of occurrence for the *to*-infinitival perfect.

²⁷ The use of BE where it is followed by a *to*-infinitival complement, being idiomatic, is often referred to as “BE *to*”. However, as we are treating all the verbs under consideration here as catenatives taking *to*-infinitival complements, for consistency we simply refer to BE. A similar point applies to the idiomatic use of HAVE (“HAVE *to*”) discussed in Section 4.3.3.

4.3.3 HAVE

HAVE can also be used to express deontic necessity, but this kind of meaning is likely to be infrequent in connection with the anteriority expressed by the perfect – though not impossible, as shown by examples like (38), which describes a requirement relating to anterior time.

- (38) *To qualify, firms **had** to have been independent and privately held through 1997.*
(1998:NEWS CSMonitor)

Many of the examples in our data are somewhat hard to categorise in terms of the type of necessity expressed. Example (39) below is perhaps best categorised as epistemic: it is from a fictional passage which appears to be taking “his” viewpoint, so we could take it as reporting a judgement on his part. Example (40) could be said to involve dynamic necessity, a situational requirement.

- (39) *She **had** to have heard him, but she didn't move.*
(1996:FIC SouthernRev)
- (40) *Besides, in order to be blackmailed one **has** to have done something disgraceful, and I know Harold didn't.*
(1941:FIC MurderMakesUsGay)

We examined the proportion of *to*-infinitival clauses that contained a perfect when preceded by HAVE, and found that it *increases* significantly from the 1950s to the 2000s. Overall, HAVE followed by a *to*-infinitival clause increases per million words over the COHA period, but the increase tails off in the last 50 years. The increase in the proportion with a perfect may be due to a rise in epistemic and dynamic uses at the expense of deontic uses; further examination of the data would be needed to test this.

4.4 The “prospective meaning” group

The final group to be considered is the “prospective meaning” group, which accounts for a far smaller share of the data than the other groups. This group comprises the four verbs MEAN (in the sense “intend”), WISH, EXPECT, and HOPE, as exemplified in (41) to (44).

- (41) *We **meant** to have sown a little barley to-day, but the ground is too dry [...]*
(1848:FIC Stranger)
- (42) *I **wished** to have purchased a mule, but, though I offered thirty pounds for a sorry one, I could not obtain her [...]*
(1843:NF BibleInSpainJourneys)
- (43) *“I **expected** to have been here in time, but these trains are never to be depended on.”*
(1871:FIC LucyRaymond)
- (44) *“[...] I **hoped** to have hidden from you what I have suffered. But it is too late now.”*
(1851:MAG Harpers)

With these verbs, the time of the infinitival clause situation is, in the absence of perfect marking, posterior to (later than) that of the matrix clause (e.g. *I mean to tell him the truth; He hopes to get a pay rise*). In the examples, the temporal relationship is also one of posteriority, and the perfect in the subordinate clause expresses counterfactual meaning (e.g. *We did not sow barley; I was not able to purchase a mule*). Such examples are frequent in the earlier data.

In each of the above, we would be more likely to find a non-perfect in present-day English: *We meant to sow a little barley today; I expected to be here in time*, and so on. The decline in counterfactual usages of the infinitival perfect may have been influenced by prescriptivism: such usages were condemned in the eighteenth and nineteenth centuries on the basis that the Latin perfect infinitive expressed only anteriority (Molenccki 2003).²⁸ However, we do find some recent examples of the perfect infinitive with “prospective” verbs in our data, such as (45).

- (45) *Of those having children in the preceding five years, 35 percent said they had not **wished** to have gotten pregnant when they did [...]*
(1994:MAG Americas)

These examples involve the catenative verb occurring in a past tense form, or in a past participle form as complement of perfect HAVE (example (45)). They rep-

²⁸ Molenccki (2003) notes that counterfactuality was in fact the earliest meaning of the English perfect infinitive: it was only after a few hundred years, around 1380, that it began to be used for non-counterfactual anteriority.

resent the simple catenative structure, with ordinary subject. The four verbs can also occur in the complex catenative structure (with the intervening noun phrase introduced by *for* in some instances, e.g. *I was hoping for him to change his mind*), but this is uncommon in our data except in the case of EXPECT.

Although examples with present tense catenatives are possible, the data for present tense verbs in this group turn out, unusually, to include a considerable number of irrelevant examples, often involving the kind of structure seen in (46).

(46) “*Well, then, ask me the question that you **wish** to have answered.*”
(1876:FIC Ishmael InDepths)

Here the string *to have answered* is not a genuine instance of the perfect – an unrelativised equivalent would be *you wish to have the question answered*.

The overall pattern of change against the “present perfect + past” baseline is a consistent fall, matching a logistic decline of more than 90 % over the period ($r^2 = 0.8650$). This pattern can be glimpsed at the bottom of Figure 9. This fall mirrors the global decline of the perfect of some 80 %.

When considering alternation patterns we decided to focus on past-marked prospective verbs only (i.e. those in past tense form, or in past participle form following perfect HAVE).²⁹ This enabled us to exclude the “noisy” data found for the present tense forms, and to focus on examples like (41)–(45) above, since we have good reason to suspect that the use of the perfect has declined in such contexts over time (given the contrasts between earlier and present-day English noted above). We therefore looked at the proportion of *to*-infinitivals which contained a perfect in the contexts following a past-marked catenative verb. The results, shown in Figure 16, show a decline in this proportion, supporting the hypothesis that speakers’ inclination to use a perfect in these contexts has declined during this period. Given this choice, the decline is even sharper: comparing points on the logistic curve, by 2000 the proportion of perfect forms is only 3 % of the same figure in 1820, or a decline of 97 %.

The fall matches a logistic regression line closely ($r^2 = 0.9117$), indicating that these verbs are being retained but writers are increasingly choosing to use the simple infinitive in preference to the perfect.

²⁹ We used the search string “-[be] wished|hoped|expected|meant” for the context, which includes past tense and past participle forms (identical for these verbs), but excludes a preceding passive BE by using the “not” symbol.

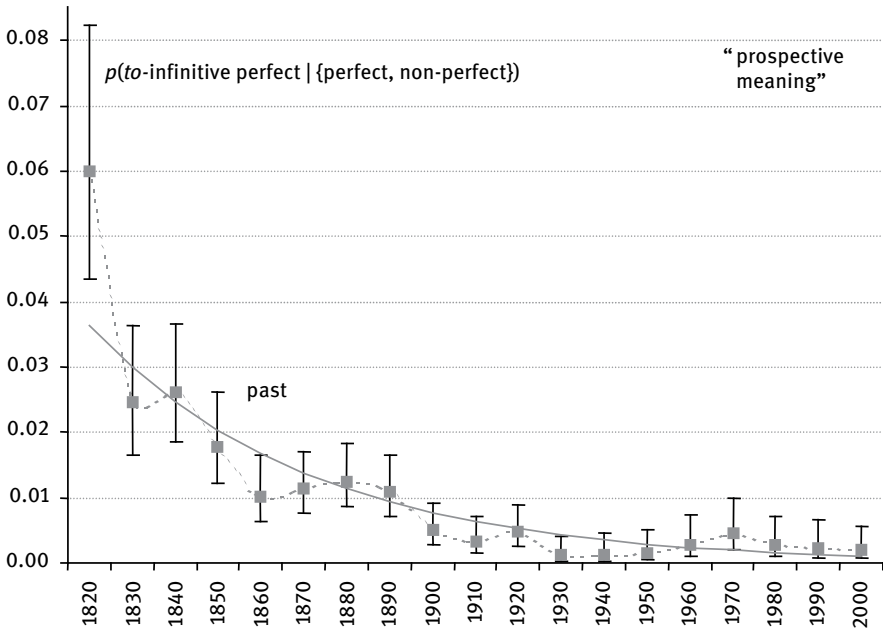


Figure 16: Decline in past-marked “prospective meaning” verbs, evaluated as an alternation between the *to*-infinitival perfect and *to*-infinitive, COHA.

4.5 The verb LIKE

The verb *LIKE* ranks tenth in our top 30 list. Two distinct uses are found in our data. The first use, illustrated in (47), is not found in present-day standard English.

(47) *But my father came in so close upon me, he **liked** to have cut his foot with the pieces.*

(1833:FIC PoemsProseWritings)

This example appears to mean “he nearly cut his foot, he seemed very likely to cut his foot”, with an implication of non-actualisation (he did not cut his foot), and a raised subject. The OED lists similar examples under *like*, *v.*², 2b: “To look like or be near *to* doing (something) or *to* being treated (in a specified manner). Now *vulgar* and *dial.* (U.S.)”. Variant forms are found in the data in reference to past time, in addition to *liked*: *like*, *had like*, *had liked*, as in (48) to (50).

- (48) [...] *he **like** to have paid dear for his conduct* [...] (1827:FIC TennesseanANovel)
- (49) *These exertions had **like** to have been fatally injurious to Constantia.* (1827:FIC Novels)
- (50) *“I’ve something to tell you, which I had **liked** to have forgotten!”* (1856:FIC DredATaleGreat)

Of 50 instances of *had like(d)* in the corpus,³⁰ none occur after 1908, but we do find some later instances of *like(d)*, e.g. *I like to have died when she asked me that* (2000:FI SouthernRev). If we consider all variants together, only about a fifth of the examples occur in the twentieth century. This use of LIKE is clearly a context of potential occurrence for the *to*-infinitival perfect which has declined over the COHA time period.

The second use of LIKE in the data typically involves a preceding *would* or *should*. This use is normal in present-day standard English, but is subject to variation in the occurrence of the perfect at both matrix and subordinate level. Compare (51) to (53).

- (51) *He would **like** to have been a KGB agent* [...] (1987:NEWS Chicago)
- (52) *“[...] I would have **liked** to have seen you one last time,” I said.* (2007:FIC BloomsdayDead)
- (53) *“Now, honey,” he began uneasily, “do you think Steve would have **liked** to see you acting this way?”* (1970:FIC PickUpSticks)

The last example is not from our data, but illustrates the possibility of using a perfect in the higher clause and not the lower. The “pastness” expressed by the perfect applies semantically to the higher clause, as reflected in the structure used in (53). However, the perfect is sometimes placed in the lower clause instead (as in (51)) or in both clauses (as in (52), where we might perhaps think of the lower perfect as expressing counterfactuality). Given this variation, it is worth

³⁰ Some of these instances have *like* tagged as an adjective, so are not included in our collocates data. This applies to (49), where *like* is indeed arguably an adjective rather than a verb. The OED mentions the use of *had like to* under *like*, *adj.* 9b.

testing whether usage has changed over time to favour a non-perfect in the lower clause. The results, shown in Figure 17, confirm that this is the case, although there is still variation.³¹

The confidence intervals are wide due to the small amount of data in the overall set, particularly in the early decades. We can see that the non-perfect becomes the dominant form around 1850, and its proportion continues to climb until around 1910. The peak in 1950 may be a text-sampling “blip”.

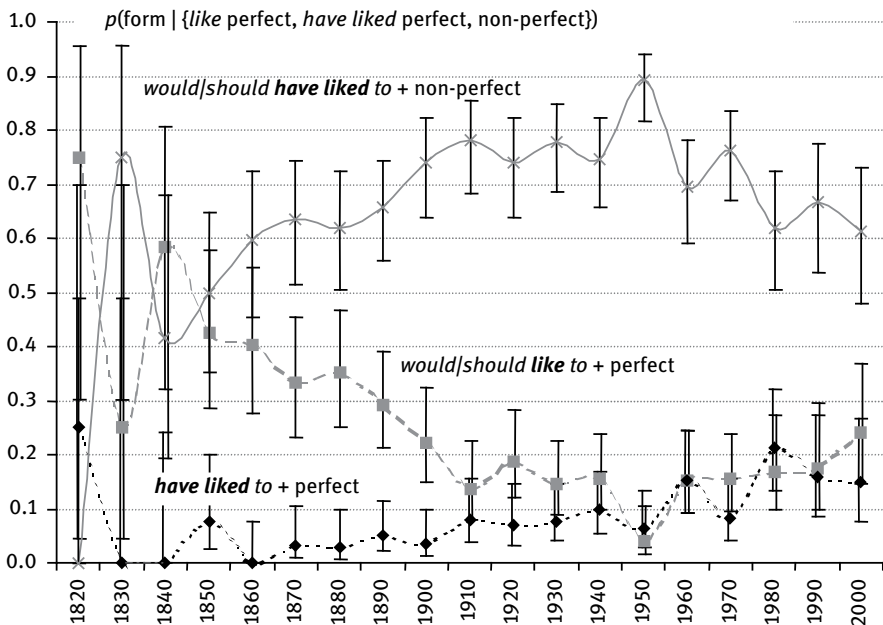


Figure 17: Variation in alternation of the perfect with *would/should* + LIKE, in COHA, by decade.

³¹ We combined the results of two search strings to obtain figures for the variant with a non-perfect in the lower clause: “would|should have liked to [v?i*] -[v?n*]”, which excludes an infinitive followed by a past participle, and “would|should have liked to be [v?n*]”, to include any passives excluded by the first search.

4.6 The verb REMEMBER

The verb REMEMBER ranks 14th in our top 30 list. Although it has a cognitive meaning, we treat it separately from the “cognition and saying” group because of its different behaviour.

There are a few examples in our data of REMEMBER in the complex catenative construction, such as the passive in (54), which has a raised subject.

- (54) *Elizabeth Scott Hardin was **remembered** to have hidden in a cave with her children [...]*
(2003:NF WhereIWasFrom)

However, the vast majority of examples in our data belong to the simple catenative construction, with an ordinary subject, as in (55) and (56).

- (55) *I **remember** to have felt some one pull me by the hair, before I was utterly senseless.*
(1835:FIC InfidelFallMexico)
- (56) *He **remembered** to have seen a photograph of this celebrity [...]*
(1911:FIC Tante)

These last two examples differ from present-day English usage, where such “retrospective” meanings are expressed with a gerund-participial complement, often but not always without the perfect, e.g. *He remembered {seeing/having seen} a photograph of this celebrity*. The *to*-infinitival complement in the simple construction is typically used in a different way in present-day English, as in *I remembered to lock the door* (meaning that I remembered that I should lock the door and did so) – a “prospective” meaning which does not lend itself to an expression of anteriority with the perfect. This contrast between the “retrospective” use of the gerund-participial complement and the “prospective” use of the infinitival complement with REMEMBER is well known by grammarians of present-day English (see, e.g., Quirk et al. 1985: 1193).

The rate of REMEMBER followed by a *to*-infinitival perfect shows a clear decline in COHA (whether per-million-words or against our verb baseline), falling to almost zero by the 1940s. The most recent examples of the kind shown in (55) and (56) above occur in the 1930s (four examples) and 1940s (one example) – apart from an example from a biography published in 1997 which involves quoted speech attributed to a man born in 1782.

Plotting the *to*-infinitival perfect against the gerund-participial³² in the context of a preceding lemma REMEMBER, we can show a temporal pattern of replacement (see Figure 18). By the 2000s, we find 707 instances with the gerund-participial, as against only one instance with the *to*-infinitival perfect, which in fact is one of the very few examples of the complex catenative construction in our data (and thus not a genuine alternant in any case). The decline shows a close match to a logistic curve, mirroring in microcosm the overall decline of the perfect (cf. Figure 3).

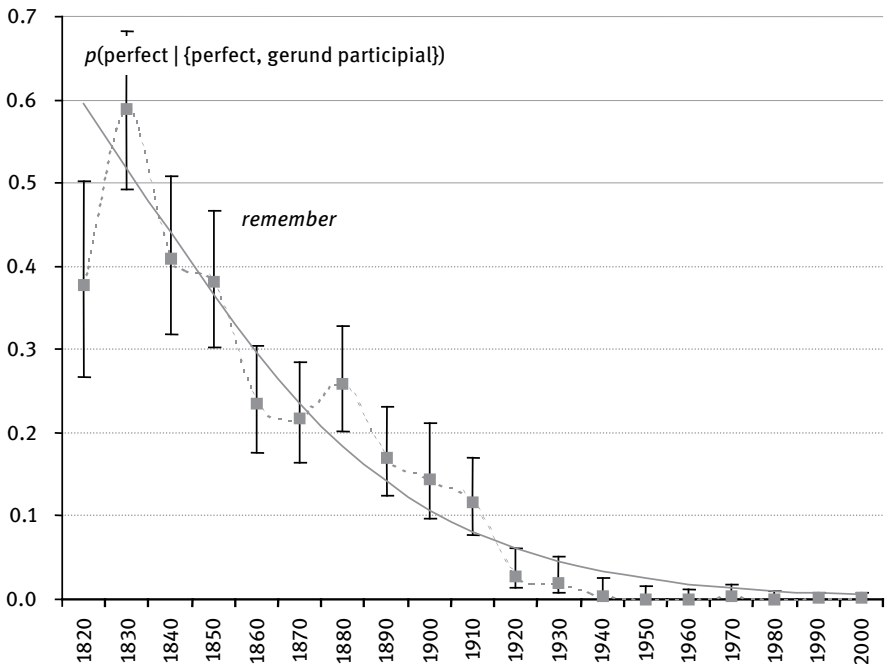


Figure 18: The decline of the perfect *to*-infinitive after REMEMBER, alternation against the gerund-participial.

Other scholars have reported on the changing complementation patterns of REMEMBER. For example, Mair (2006) presents data from the OED quotation base for 1700 to 2000, having found available corpora unsuitable for the purpose (for example, ARCHER contained too few instances of the relevant constructions). On

³² The gerund-participial was searched for using “[v?g]”, which includes both perfect and non-perfect instances.

the one hand, Mair traces the origin of the gerund-participial complement back to the late eighteenth century, citing two quotations from the 1770s. On the other, he notes two authentic examples of the “retrospective” use of the *to*-infinitival perfect occurring as recently as the 1960s and 1970s. It is hard to resist quoting one of these: *It is the first play in which I remember to have encountered an actor with a jockstrap which squeaks when pushed* (*Daily Telegraph*, 7 March 1969).

4.7 The remaining verbs

The remaining ungrouped verbs are PROVE, HAPPEN, and PRETEND, ranking at 17, 18, and 20 respectively in our top 30 list. PROVE is ungrouped because it occurs with two different senses in our data, illustrated in (57) and (58).

(57) *The Moscow “failure” might yet **prove** to have been a successful first step.*
(1977:MAG Time)

(58) *The defendant was **proved** to have struck a horse twenty-five times upon the head with a heavy club [...]*
(1873:NEWS NYT-Ed)

In (57), where PROVE has the sense of “turn out”, we have a simple catenative structure with a raised subject, as found with the “seeming” verbs. In (58), we have a passive complex catenative structure with a raised subject, as found with the cognition and saying group. The simple structure seems to be dominant in our data, with a substantial minority of the complex.

We have already mentioned that HAPPEN shares many properties with the “seeming” verbs, but as it is different in meaning and far less frequent, we have not included it in the same group. (59) and (60) are a couple of examples.

(59) *“Meta **happens** to have fallen in love with a young gentleman named Edmund Innis [...]*
(1872:FIC DoctorVandykeA)

(60) *He **happened** to have brought along a poem of his on the white wines of Sparta.*
(1970:FIC WreckageAgathon)

Finally, PRETEND occurs in our data mainly in the simple catenative construction, with an ordinary subject, as in (61) and (62).

- (61) *They **pretend** to have invented the printing press [...]*
(1841:NF PsychologyOrAView)
- (62) *I hit the floor, **pretending** to have fainted dead away.*
(2003:FIC Fudge-a-mania)

These examples seem to represent two somewhat different senses, the first being close to “claim” in meaning. The two senses are not always clearly distinct, however; a clear “claim” sense may belong to earlier usage. The “claim” sense also allows a complex construction, as in (63), one of very few passives found in the data, all from the nineteenth century.

- (63) *No aid and no countenance can the labors of the apostles be **pretended** to have received from any secular power.*
(1842:NF AnInquiryIntoFoundation)

Despite the small figures and large error bars for these individual verbs, we can identify a small decline over the entire period (1820–2000) against the past tense + present perfect baseline for the *to*-infinitival perfect for both PROVE and PRETEND. However, there is insufficient data to identify a similar trend for HAPPEN.

5 Conclusions

Our study has focused on the *to*-infinitival perfect, a subtype of the perfect that has received far less attention than the present perfect. We looked at changing use of this structure in written AmE over the period 1820–2009. Against a baseline of present perfect and past tense verbs, we found a huge overall decline of around 80 % over this time period. We chose this baseline because perfect *to*-infinitive forms correspond with past-referring contexts, whether expressed by a present perfect or by a past tense verb. Other potential generic baselines, such as words or *to*-infinitives in general, risk introducing both additional noise and skewing by other factors, such as variation in the temporal orientation of texts sampled.

In order to investigate the decline in more detail, we needed to examine the contexts in which the *to*-infinitival perfect occurs. We used ICE-GB, with its parse analysis and powerful search tools, to obtain a reliable snapshot of comparable data (albeit 1990s BrE, 60 % spoken) and thereby to ascertain the different kinds of structures we could expect to find in COHA, a much larger corpus with more limited annotation and tools. We found a large majority of examples occurring

as complements of catenative verbs. Our diversion into ICE-GB further allowed us to anticipate the distribution of simple and complex catenative constructions, including passive complex ones. It enabled us to design POS-tag query strings for COHA to obtain results, which, while not guaranteed to cover all cases, were reasonably reliable.

Use of COHA, with its 400 million words covering two centuries of written AmE, then provided ample data for a detailed diachronic study of the catenative verb collocates of the *to*-infinitival perfect. Given the exponential frequency distribution of these collocates, we opted to focus on the most numerous verbs, because changes affecting the “long tail” of 5% of tokens could not numerically explain the scale of the overall decline in the perfect.

We found that many of the “top 30” collocates express meanings in the broad areas of evidentiality and modality, and so could be thought of as making “evidentialising” and “modalising” contributions to the situation described in the infinitival complement clause. Within this broad area, however, we could identify some distinct subsets. We used criteria drawn from Huddleston and Pullum (2002) to group verb lemmas, both semantically and grammatically, taking into account their possible alternation patterns (i.e. the options available to an author as alternatives to using the perfect *to*-infinitive).

By grouping verbs in this way and studying their change over time, we found a range of different patterns. Some cases – including the “seeming” group, the “prospective” group, and the verb REMEMBER – decline according to a logistic curve: the simplest, most consistent form of decline. They therefore conform to the overall pattern of decline, although they make differing quantitative contributions to this trend.

Since the “seeming” verbs are most frequent (covering 50% of the data), their apparently regular decline might lead one to conclude that all verbs would behave like this. This is not the case. Examining the “cognition and saying” group reveals a number of different changes taking place at the same time. First KNOW declines, at the beginning of the period examined. SAY, the predominant form, holds more-or-less steady until around 1920, and then starts to fall quite substantially. REPORT and CLAIM expand in use in the late twentieth century, such that by the end of the period the overall decline is combined with a growth in the diversity of forms. Closer study, including genre comparison, would be needed to investigate whether these patterns reflect stylistic changes, or extra-linguistic, social changes of some kind. This group also revealed a striking degree of variation in their “strength of attraction” to the perfect, considered synchronically: another aspect worth further investigation.

Similarly, the “modality” group also covers three verbs which behave very differently. Most obviously, *ought* rises before falling – a pattern attributed to its

replacement by the core modal *should*. By contrast, *was/were* and *HAVE* are relatively constant in their proportion (of present perfect and past verbs) for around a century, until the 1920s, when *was/were* declines and *HAVE* increases.

The groups and verbs also show marked contrasts in terms of the alternation processes in which they participate over time. For the “seeming” group, and *SAY* (the predominant “cognition and saying” verb), we were unable to identify any clear pattern of replacement of the *to*-infinitival perfect by a particular form: the most obvious potential alternant, the finite complement clause, could not account for the declines. By contrast, for other groups and verbs, clear patterns of replacement could be identified. We saw replacement by non-perfect *to*-infinitival complements for the “prospective” group (earlier *I expected to have been here in time* vs. present-day *I expected to be here in time*), and for *LIKE* preceded by *would/should (have)*, where ongoing variation remains despite the increase in proportions of the non-perfect complement with a superordinate perfect (as in *I would have liked to go*). A different pattern again was seen with *REMEMBER*, with *to*-infinitival complements clearly being replaced over time by gerund-participial complements (earlier *I remember to have seen him there* vs. present-day *I remember seeing him there*).

Given the wide range of different patterns we have found, what can we conclude about the decline in this form, and about the way language changes in general? It may appear that the observed decline is simply the “accidental” result of various unconnected changes in the language which combine to create this effect.

However, the fact that the overall trend is progressively downwards suggests that there is indeed a decreasing general inclination to use the *to*-infinitival perfect, a trend which manifests itself in various ways across different constructions and collocational contexts. Some contexts may buck the trend because of other influences, whether linguistic or extra-linguistic.

This view is reinforced by the observation that the same form declines significantly in COHA in other grammatical contexts besides those we focused on in this study. For instance, further searches show that it also declines following adjective collocates, and in complex active catenative constructions with an intervening pronoun (excluded by our main data). Furthermore, there is a steep decline over the COHA period in another category of non-finite perfect, the gerund-participial perfect (as in *Having finished her work, she left early*). These observations about declines in related structures suggest there is some more general tendency at work.

Our findings of lexical variation within a more general overall trend fit quite well with recent strands of work on language change within a constructional approach (e.g. Hilpert 2013), which stress that grammatical change tends

to be lexically constrained and gradually diffusional in nature rather than rapid and systemic. This is certainly a fertile area for further research. Tracing such complex, lexically conditioned change processes has become much more feasible with the availability of megacorpora spanning long time periods such as COHA.

Acknowledgements

We are grateful to two anonymous reviewers for their helpful, detailed comments. We would also like to thank Seth Mehl and Martin Hilpert for their comments on drafts of the paper, and Mark Davies for his extensive answers to queries about COHA.

References

- Bowie, Jill & Bas Aarts. 2012. Change in the English infinitival perfect construction. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 200–210. Oxford: Oxford University Press.
- Bowie, Jill, Sean Wallis & Bas Aarts. 2013a. The perfect in spoken British English. In Bas Aarts, Joanne Close, Geoffrey Leech & Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 318–352. Cambridge: Cambridge University Press.
- Bowie, Jill, Sean Wallis & Bas Aarts. 2013b. Contemporary change in modal usage in spoken British English: Mapping the impact of “genre”. In Juana I. Marín Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), *English modality: Core, periphery and evidentiality*, 57–94. Berlin: Mouton de Gruyter.
- Callies, Marcus. 2013. Bare infinitival complements in Present-Day English. In Bas Aarts, Joanne Close, Geoffrey Leech & Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 239–255. Cambridge: Cambridge University Press.
- Collins, Peter. 2009. *Modals and quasi-modals in English*. Amsterdam: Rodopi.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Davies, Mark. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* 7(2). 121–157.
- Elsness, Johan. 1997. *The perfect and the preterite in contemporary and earlier English*. Berlin: Mouton de Gruyter.
- Elsness, Johan. 2014. The present perfect and the preterite in Late Modern and contemporary English: A longitudinal look. In Kristin Davidse, Caroline Gentens, Lobke Ghesquière & Lieven Vandelanotte (eds.), *Corpus interrogation and grammatical patterns*, 81–103. Amsterdam: Benjamins.
- Fanego, Teresa. 2007. Drift and the development of sentential complements in British and American English from 1700 to the present day. In Javier Pérez-Guerra, Dolores González-

- Álvarez, Jorge L. Bueno-Alonso & Esperanza Rama-Martínez (eds.), *'Of varying language and opposing creed': New insights into Late Modern English*, 161–235. Bern: Peter Lang.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2005. *A student's introduction to English grammar*. Cambridge: Cambridge University Press.
- Hundt, Marianne & Nicholas Smith. 2009. The present perfect in British and American English: Has there been any change, recently? *ICAME Journal* 33. 45–63.
- Mair, Christian. 2006. Nonfinite complement clauses in the nineteenth century: The case of *remember*. In Merja Kytö, Mats Rydén & Erik Smitterberg (eds.), *Nineteenth-century English: Stability and change*, 215–228. Cambridge: Cambridge University Press.
- Moléncki, Rafał. 2003. Proscriptive prescriptivists: On the loss of the 'pleonastic' perfect infinitive in counterfactual constructions in Late Modern English. In Marina Dossena & Charles Jones (eds.), *Insights into Late Modern English*, 175–196. Bern: Peter Lang.
- Nelson, Gerald, Sean Wallis & Bas Aarts. 2002. *Exploring natural language: Working with the British component of the International Corpus of English*. Amsterdam: Benjamins.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.
- Vosberg, Uwe. 2009. Non-finite complements. In Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*, 212–227. Cambridge: Cambridge University Press.
- Wallis, Sean. 2013. Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics* 20(3). 178–208.
- Wallis, Sean. 2014. What might a corpus of parsed spoken data tell us about language? In Ludmila Veselovská & Markéta Janebová (eds.), *Complex visibles out there. Proceedings of the Olomouc Linguistics Colloquium 2014: Language use and linguistic structure*, 641–662. Olomouc: Palacký University.
- Werner, Valentin. 2014. *The present perfect in World Englishes: Charting unity and diversity*. Bamberg: University of Bamberg Press.
- Yao, Xinyue & Peter Collins. 2012. The present perfect in world Englishes. *World Englishes* 31(3). 386–403.
- Yao, Xinyue & Peter Collins. 2013. Recent change in non-present perfect constructions in British and American English. *Corpora* 8(1). 115–135.

Appendix A

Table of COHA sampling distribution (number of words), drawn from <http://corpus.byu.edu/coha/old> under “Composition of the corpus”. See also Figure 1. This distribution affects our results in the following two main ways: The lower total frequencies for the earlier decades tend to obtain wider confidence intervals (lower certainty). The low proportion of magazines in the first decade (compounding the absence of newspaper data) causes us to discount this data point.

Decade	Fiction	Magazines	Newspapers	NF books	Total
1810	641,164	88,316	0	451,542	1,181,022
1820	3,751,204	1,714,789	0	1,461,012	6,927,005
1830	7,590,350	3,145,575	0	3,038,062	13,773,987
1840	8,850,886	3,554,534	0	3,641,434	16,046,854
1850	9,094,346	4,220,558	0	3,178,922	16,493,826
1860	9,450,562	4,437,941	262,198	2,974,401	17,125,102
1870	10,291,968	4,452,192	1,030,560	2,835,440	18,610,160
1880	11,215,065	4,481,568	1,355,456	3,820,766	20,872,855
1890	11,212,219	4,679,486	1,383,948	3,907,730	21,183,383
1900	12,029,439	5,062,650	1,433,576	4,015,567	22,541,232
1910	11,935,701	5,694,710	1,489,942	3,534,899	22,655,252
1920	12,539,681	5,841,678	3,552,699	3,698,353	25,632,411
1930	11,876,996	5,910,095	3,545,527	3,080,629	24,413,247
1940	11,946,743	5,644,216	3,497,509	3,056,010	24,144,478
1950	11,986,437	5,796,823	3,522,545	3,092,375	24,398,180
1960	11,578,880	5,803,276	3,404,244	3,141,582	23,927,982
1970	11,626,911	5,755,537	3,383,924	3,002,933	23,769,305
1980	12,152,603	5,804,320	4,113,254	3,108,775	25,178,952
1990	13,272,162	7,440,305	4,060,570	3,104,303	27,877,340
2000	14,590,078	7,678,830	4,088,704	3,121,839	29,479,451
Total	207,633,395	97,207,399	40,124,656	61,266,574	406,232,024

Appendix B

COHA search for left-hand verb collocates of the *to*-infinitival perfect (screenshot fragment).

The screenshot shows the COHA search interface with the following settings:

- DISPLAY:** LIST (selected), CHART, KWIC, COMPARE
- SEARCH STRING:**
 - WORD(S): to have [v?n*]
 - COLLOCATES: [v*] 2 0
 - POS LIST: (empty)
 - RANDOM: (button)
 - SEARCH: (button)
 - RESET: (button)
- SECTIONS:** SHOW (checked)

1	--IGNORE--	2	--IGNORE--
2000		2000	
1990		1990	
1980		1980	
1970		1970	
1960		1960	
1950		1950	
- SORTING AND LIMITS:**
 - SORTING: FREQUENCY
 - MINIMUM: FREQUENCY 5
- HIDE OPTIONS:**
 - # HITS: (empty)
 - FREQ: 500
 - KWIC: 100
 - GROUP BY: LEMMAS
 - DISPLAY: RAW FREQ
 - SAVE LISTS: NO