Variation and change in the productivity of BE *going to* V in the *Corpus of Historical American English*, 1810–2009[1]

TANJA SÄILY
University of Helsinki

FLORENT PEREK
University of Birmingham

JUKKA SUOMELA
Aalto University

ABSTRACT

This paper focuses on the poster child of grammaticalization, BE *going to* V. First expressing 'motion with intention', in Early Modern English the construction came to signify 'motionless intention'. The grammaticalization process continued in Late Modern English with subjectification, so that 'intention' was gradually replaced by 'prediction'. We study the process from Late Modern to Present-Day English in the 200-million-word fiction section of the *Corpus of Historical American English* (COHA), 1810–2009, for which we have gender metadata on the authors. We focus on the productivity of the construction by comparing type frequencies, i.e., the number of different verbs following BE *going to*. Our research questions are how the grammaticalization is reflected in the productivity of the construction, and whether the social factor of gender played a role in the process. We study the internal factors of mental verbs, inanimate subjects and passive voice; to this end, we use robust statistical methods to compare type frequencies and proportions of types over time. We also investigate the semantics of the verb types by drawing on techniques from distributional semantics. Our wider aim is to enrich the cognitively oriented theory of Construction Grammar with insights from historical sociolinguistics.

Keywords: grammaticalization, productivity, distributional semantics, future constructions, historical sociolinguistics

1 INTRODUCTION

Our paper focuses on the poster child of grammaticalization, BE *going to* V (e.g. Bybee & Pagliuca 1987; Danchev & Kytö 1994; Hopper & Traugott 2003; Hilpert 2008; Traugott & Trousdale 2013; Budts & Petré 2016; Petré & Van de Velde 2018). First expressing 'motion with intention', as in *I'm going to the market to buy bananas*, in Early Modern English the construction came to signify 'motionless intention', as in *I'm going to read your work tomorrow*. The grammaticalization process continued in Late Modern English with subjectification, so that 'intention' was gradually replaced by 'prediction', as in *There's going to be some serious trouble here* (Budts & Petré 2016). Wu *et al.* (2016) conducted a quantitative analysis of the grammaticalization of BE *going to* V in the *Corpus of Historical American English* (COHA), 1810–2009. They found that the frequency of the construction continued to increase in COHA from Late Modern to Present-Day English, also in terms of the use of mental verbs, inanimate subjects and passive voice, all of which can be linked to the 'prediction' sense.

We study this juncture in the 200-million-word fiction section of COHA, using gender metadata developed by Öhman *et al.* (2019). Whereas Wu *et al.* (2016) analysed the grammaticalization process in terms of token frequency, we focus on the productivity of the construction by comparing type frequencies, i.e., the number of different verbs following BE *going to*. Our research questions are how the grammaticalization is reflected in the productivity of the construction, and whether the social factor of gender played a role in the process. Following Wu *et al.* (2016), we study the internal factors of mental verbs, inanimate subjects and passive voice; we also investigate the semantics of the verb types by drawing on techniques from distributional semantics (Perek 2016), which allows us to measure the semantic spread of the construction at different points in time. To compare type frequencies and proportions of types over time and across factors, we use robust statistical methods building upon the work of Rodríguez-Puente *et al.* (2022) and Säily *et al.* (2024).

Although the research reported in this study is not dependent on any particular theoretical framework, our wider aim is to enrich the cognitively oriented theory of Diachronic Construction Grammar (DCxG; e.g. Traugott & Trousdale 2013) with insights from historical sociolinguistics. Studies in both present-day and historical sociolinguistics have found the social category of gender to be key in language change. In particular, it is often women who tend to lead language change (e.g. Labov 2001: 262; Nevalainen & Raumolin-Brunberg 2003: 131). Previous sociolinguistic work on variation and change in

morphological productivity, however, has tended to show a male advantage (e.g. Säily 2014), the reasons for which are as yet unclear. It is therefore of interest to study the patterning of gender in the productivity of syntactic constructions as well, particularly in combination with language-internal factors (cf. Säily *et al.* 2024).

The rest of this paper is organized as follows. Section 2 surveys previous research on the construction, while section 3 discusses our material. Section 4 presents our methods and analysis, and section 5 concludes the paper with a discussion of the results and their implications for future research.

2 BACKGROUND

Movement is one of the main sources of future constructions cross-linguistically (e.g. Bybee & Pagliuca 1987: 110–1). Bybee & Pagliuca (1987: 116–18) propose that the meaning of BE *going to* V grammaticalized from concrete movement to figurative movement (including a sense of intention) and ultimately to prediction, similarly to many other future constructions (Bybee *et al.* 1991: 32). Hopper & Traugott (2003: 2–3) argue that the starting point of the process was a directional use that already included a sense of intention or purpose. Budts & Petré (2016: 2, 16) provide examples of the stages of grammaticalization of the construction, summarized in (1)–(3) below.

(1) I'm going to the market to buy bananas. [motion with intention]
(2) I'm going to read your work tomorrow. [motionless intention]
(3) There is going to be some serious trouble here [prediction]

Previous research indicates that the 'motionless intention' sense arose in Early Modern English by the late seventeenth century (e.g. Danchev & Kytö 1994; Núñez Pertejo 1999), while the 'prediction' sense developed in Late Modern to Present-Day English (e.g. Budts & Petré 2016; Wu *et al.* 2016). The stages overlap, and in fact all of the senses still coexist today, which is not the case with all movement-based future constructions (Traugott 2012: 241). Rather than absolute distinctions, the change therefore chiefly operates on the relative proportions of the different senses, as also noted by Danchev & Kytö (1994: 68–9). In DCxG terms, the emergence of the motionless intention sense corresponds to an instance of constructionalization (Traugott & Trousdale 2013), i.e. the creation of a new form-meaning pair in the network of constructions and its conventionalization across the speech community. In other words, the BE *going to* V pattern comes to be directly associated with the

meaning of intention found in instances of *go* as a motion verb such as (1), and thus loses the entailment of motion. As discussed by Traugott & Trousdale (2013), constructionalization is typically followed by constructional changes, i.e. changes in the form of the construction (e.g. reduction), its meaning (in particular semantic bleaching), and its productivity.

The semantic change of the BE *going to* V construction from intention to prediction can be regarded as a process of subjectification, by which constructions begin to be used to express the speaker's attitudes and beliefs (Traugott 2010: 33). The subjectification of BE *going to* V in Late Modern English was analysed in a relatively large-scale empirical study by Budts & Petré (2016), who found in the *Corpus of Late Modern English Texts* a number of indicators of change from the late eighteenth to the early nineteenth century. These included a decrease in first-person subjects and imminent future contexts (*I am going to entrust you with a secret*) as well as an increase in non-agentive subjects and infinitival complements expressing events of low agentivity (*It's going to rain*; our example from COHA, 1811).

Wu *et al.* (2016) also examined the later development of the construction, focusing on American English in COHA from the nineteenth century onwards. They investigated the frequencies of mental verbs, the inanimate subject *it* and passive voice in the construction over time, as in (4)–(6) (our examples). All of these can be linked to the change from intention to prediction, as 'this usage decreases the intention and purposiveness of the subject as well as the directionality of the verb *go*' (Wu *et al.* 2016: 321). Wu *et al.* (2016) found that the frequencies of these contexts increased throughout the time period covered by the corpus, from 1810 to 2009, which to us indicates continued grammaticalization, even if Wu *et al.* (2016: 324) consider the construction to have been 'fully grammaticalized' in the late seventeenth century, as determined by the first occurrence of mental verbs in the construction in their Google Books data.[2]

(4) You're going to feel very foolish (COHA, 1932) [mental verb]
(5) It's going to rain (COHA, 1811) [inanimate subject *it*]
(6) Father Paul was going to be cheated of his share (COHA, 1946) [passive voice]

---

[2] Following Hopper & Traugott (2003: 6), we view grammaticalization as a cline, also from the perspective of frequency: 'the more frequently a form occurs in texts, the more grammatical it is assumed to be' (Hopper & Traugott 2003: 106).

It has often been noted that changes in the meaning of a grammaticalizing construction are typically correlated with what Himmelmann (2004: 32) terms 'host-class expansion', i.e. an increase in the range of lexical items attested in the slots of the construction. In a DCxG approach, these changes may directly follow from semantic change in the construction: as the meaning of a construction becomes more schematic, a wider set of lexical items become compatible with it and may thus be used in the construction (Perek 2018, 2020). An example of this for the BE *going to* V construction is the case of mental verbs investigated by Wu *et al.* (2016), as already mentioned above: mental predicates such as *forget* and *amuse* do not easily lend themselves to the original 'motion with intention' meaning of the BE *going to* V pattern, or to the newly grammaticalized meaning 'motionless intention'. Hence, their occurrence in the construction speaks to the increased availability of the more abstract prediction sense. Importantly, host-class expansion does not typically occur abruptly but is rather gradual and probabilistic (Traugott & Trousdale 2013: 114–15), hence it can be documented in diachronic corpora. Mental verbs are but one example of host-class expansion in BE *going to* V; more broadly, Hilpert (2008: 119–21) reports from a diachronic distinctive collexeme analysis of the construction in the CLMET corpus that verbs with high agentivity such as *fight* and *speak* are most prominent in the earlier periods of the corpus (i.e. 18th to mid 19th century), while more schematic verbs like *be*, *do*, *have* and verbs referring to non-intentional events like *die* and *happen* are most distinctive of the last period (see also Budts & Petré 2016: 20–1). However, while these studies do confirm the grammaticalization path of BE *going to* V through some of its distributional changes, a more detailed description of host-class expansion in the history of construction is lacking.

We build upon the work of Budts & Petré (2016) and Wu *et al.* (2016) by considering two hitherto unexamined perspectives on the process in Late Modern and Present-Day English. First, we focus on changes in the productivity of the construction, in terms of overall type frequencies of the verb slot, the proportions of types occurring in the three contexts exemplified in (4)–(6) above, as well as host-class expansion, in terms of new semantic types of verbs joining the distribution of the construction.[3] Second, we combine our constructional approach with historical sociolinguistics by considering the influence of gender on the change. For this purpose, we utilize the fiction section of COHA, which has identifiable authors.

---

[3] For discussion of how measures of this kind relate to potential mental representations of language users, see e.g. Baayen (1993) and Perek (2016: 163–164).

3 MATERIAL

We use the first version of COHA (Davies 2010–), which contains approximately 400 million words from the period 1810–2009. Aside from it being the version to which our institutions had access, the choice of the first version was motivated by comparability with previous research (Wu *et al.* 2016) as well as by the availability of fiction-specific genre and gender metadata for this version (see below). The instances of the construction were retrieved from the downloadable version of the corpus licensed by the University of Helsinki, which makes the data more easily accessible but has the downside that 5% of the tokens, or ten consecutive words in every 200 words, have been replaced with @ signs for copyright reasons.[4] The corpus nevertheless remains usable for linguistic research, and its large size enables the study of lower-frequency constructions over time.

COHA is divided into four genres: fiction (c. 50% of the data), magazines, newspapers and other non-fiction. As fiction is usually written by identifiable authors, it was used as the basis for our sociolinguistic study. Furthermore, fiction represents informal language use with elements of speech-like interaction (Culpeper & Kytö 2010: 16–8), where many kinds of language change are expected to show up earlier than in more formal genres. The fiction section of COHA comes with metadata on the names of the authors, although these are not in a completely consistent format. This metadata was used by Öhman *et al.* (2019) to generate gender information for the authors.[5] The identification was based on extracting the first names of the authors, which were then compared with historical parish records as well as modern name lists using a machine learning approach to label the author as male or female, with uncertain cases labelled as unknown. This resulted in a highly accurate classification which enabled Öhman *et al.* (2019) to discover consistent gender patterns in the use of the indefinite pronouns -*body* and -*one* over time. For the present study, we did some further spot-checking and manually corrected the coding of the ambiguous name Robin, and one case in which the author was a husband-wife pair, for which we set the gender to unknown. The size of the fiction dataset for which gender is known is c. 186 million words out of a total of 197 million words.

There are, however, some issues with the balance of the fiction section of COHA, in that different kinds of fiction are unevenly distributed over time. While novels form the majority of the corpus, it also contains for instance short

---

[4] https://www.corpusdata.org/limitations.asp

[5] We thank Emily Öhman for providing us with the enriched metadata file.

stories, drama and movie scripts, the proportion of which increases in the twentieth century. Säily & Vartiainen (forthcoming) noticed that this skewed their results on the change from *much -ed* to *very -ed*, so that there seemed to be a gender difference in the latter part of the corpus which disappeared when the dataset was restricted to novels alone. Following their lead, we have decided to restrict our dataset to novels only. The size of this dataset by gender and decade is shown in table 1. Note that for most of our analyses we use a sliding window of 40 years with 10-year increments: the 40-year window ensures that we have enough data for each window position, while moving it by such a small increment enables us to more accurately identify periods of significant variation and change.

Table 1. *Size of the dataset by decade and author gender. Total word counts include authors of unknown gender*

Decade	Word count		
	Total	Female	Male
1810–1819	492,472	63,328	419,027
1820–1829	3,281,590	820,014	2,413,139
1830–1839	6,812,445	956,371	5,847,472
1840–1849	8,224,846	1,023,672	6,953,209
1850–1859	8,488,714	2,842,557	5,414,286
1860–1869	8,676,217	3,057,175	5,319,779
1870–1879	9,530,930	3,547,409	5,949,840
1880–1889	10,273,121	3,347,553	6,679,136
1890–1899	10,554,419	3,033,890	7,232,192
1900–1909	10,580,235	3,557,375	6,797,700
1910–1919	10,118,126	3,650,398	6,257,330
1920–1929	10,321,559	2,908,490	7,088,307
1930–1939	7,205,245	2,194,186	4,187,759
1940–1949	7,689,242	2,156,918	5,164,514
1950–1959	6,864,605	1,495,436	5,369,169
1960–1969	6,517,254	1,404,627	5,065,525
1970–1979	8,001,898	1,763,572	5,798,294

1980–1989	8,282,047	2,743,141	5,353,102
1990–1999	3,394,376	1,628,038	1,614,090
2000–2009	9,043,070	4,406,704	3,641,736
Total	154,352,411	46,600,854	102,565,606

We acknowledge that the use of this dataset to analyse gender differences in language change is somewhat problematic. The dataset is not a socially representative or balanced corpus of American men and women's language use in 1810–2009; rather, it represents the language use of the highly selective group of published authors of fiction included in COHA, the sampling of which has been influenced by such issues as availability and copyright. Moreover, the language use in novels varies not only by the gender of the author, but also by the gender of the characters as well as by subgenre, on which we have no metadata. Nevertheless, as novels tend to be informal and speech-like, and there are plenty of them in the corpus, this is arguably the best dataset currently available for the gender-based analysis of this 200-year stretch of American English. Encouraged by the results of Öhman *et al.* (2019), we embrace Labov's (1994: 11) definition of historical linguistics as 'the art of making the best use of bad data' and proceed with our analysis, while keeping the limitations of the dataset in mind and returning to them in section 5.

Our data retrieval and cleanup process was as follows. We used the 'database form' of COHA,[6] including the 'db', 'lexicon' and 'sources' (metadata) files. We combined the metadata from the database with fiction-specific genre metadata provided in the 'cohaTexts.xls' file downloaded from the English-Corpora.org site, as well as with the gender metadata produced by Öhman *et al.* (2019). We did some manual cleanup of the metadata as described above and turned it into a machine-readable (JSON) form.

To find relevant instances of the construction, we automatically extracted data from the COHA database files using a Rust program. The search query was `VB*, "going", "to", V?I*`. We also searched for the *gonna* form (`"gon", "na", *`), which emerges in the 1920s in the novels of our corpus and represents the final stage of grammaticalization of BE *going to* V, in which the complex auxiliary has been reanalysed to a single morpheme (Hopper & Traugott 2003: 68), but the dataset ended up being too small for our

---

[6] https://www.corpusdata.org/database.asp

study.[7] We then cleaned up the results for the fiction section manually in Excel. In total, 1,861 BE *going to* V types were checked, and verb types like *market*, which was actually a noun (as in *I was going to market to get some of the best fruit*), were removed. In a later step, instances of the *be*-passive (defined as `VB*, V?N`) and *get*-passive (`"get", V?N`) were classified and corrected so that the verb that was counted was the lexical verb, e.g. *ask* rather than *be* in *was going to be asked*. Other classifications added to the data were mental verbs (using a list from Halliday & Matthiessen (2014: 256–7), with ambiguous types like *strike* removed) and the inanimate subject *it* (defined as `"it"` at the end of the context preceding the node). Our software library for querying COHA, as well as the dataset with our classifications and metadata in a JSON format, are freely available for download.[8]

4 METHODS AND ANALYSIS

4.1 *Type frequencies*

4.1.1 *Challenges in analysing type frequencies*

To study how the productivity of the BE *going to* V construction has changed over time, we analyse type frequencies: we count how many different verbs follow BE *going to* in each time period. However, here we face three main challenges.

First, we have vastly different amounts of text from different time periods, and different amounts of text from men and women (see table 1). Type frequencies cannot be directly compared across corpora of different sizes: for example, if a 1-million-word corpus has 100 different verbs in BE *going to* V constructions, it is not at all obvious how many types we would expect in a 10-million-word corpus. Normalization is not an option because type frequencies grow nonlinearly with corpus size (Säily & Suomela 2009: 96–97).

Second, we would like to be able to make a distinction between the BE *going to* V construction being used more frequently vs. it being used in a more diverse manner. Even if two texts have the same number of words, one of them might use BE *going to* V more frequently but in a repetitive manner, while

---

[7] While we decided to keep the two forms separate for this study, the results would not change significantly if the type frequencies of BE *going to* V and *gonna* V were combined in the analysis, as the latter dataset is so small.

[8] https://github.com/suomela/coha-filter; link to dataset to be added after gaining permission from Emily Öhman to share the gender metadata.

another one could use it less frequently but in a productive manner. Merely counting the number of types will miss this distinction.

Third, even if we had the same amount of data and the same token frequency, and we observed differences in the type frequencies, we would still need to be able to assess whether the differences are statistically significant.

4.1.2 *Methods for analysing type frequencies*

We address these challenges in our study by building upon the methodology developed by Säily & Suomela (2009, 2017), Rodríguez-Puente *et al.* (2022), and Säily *et al.* (2024). The methods build on the idea of resampling (more precisely, permutation testing): to assess whether women use BE *going to* V in a significantly productive manner, we compare the number of types in the subcorpus that consists of women's texts with the number of types in a randomly constructed subcorpus of the same size. For example, if we have 1 million words written by women and we have 100 different types in their texts, we can construct a random subcorpus with 1 million words and see if we get as many types there. We can repeat this process a large number of times, and this way conclude whether 100 types is an unusually high number in this corpus for a subcorpus of such length. We can directly assess the statistical significance of this finding by counting which fraction of random subcorpora have such a high number of types. For example, if fewer than 0.1% of random subcorpora have at least as high a number of types as the texts written by women, then we can conclude that gender indeed is significant with $p < 0.001$ for a one-tailed test or with $p < 0.002$ for a two-tailed test; in the two-tailed version the null hypothesis is simply that gender is unrelated to type diversity. The numbers we report here are for a one-tailed test.

In a similar manner, we can construct a subcorpus of texts written in a certain time period, and assess whether the number of types there is unusually high or low in comparison with a random subcorpus of a similar size. Here random sampling is done at the level of texts (as opposed to sampling random words), to ensure that any single text cannot have a large impact on the results, and to ensure that the randomly constructed corpus contains meaningful textual units, similar to the original corpus.

A direct application of the above idea already addresses the first and third issues listed in 4.1.1 above: we are comparing subcorpora of the same size, and by constructing a large number of random subcorpora we can assess statistical significance. However, the technique can also be adapted to different measures of corpus size: instead of sampling random subcorpora with a given

number of running words, we can sample random subcorpora with a given token frequency of the target construction. This enables us to address the second issue.

4.1.3 *The* types*3 software*

For this work, we have developed a computer program that we call *types*3, which makes it possible to explore trends in type frequencies and to assess the statistical significance of the trends (see figure 1). The software loads a JSON file that describes the composition of the corpus – in our case, the JSON file lists all distinct texts of the corpus, and for each text it contains text-level metadata (e.g. the gender of the author) and a list of all instances of BE *going to* V in the text. For each instance we have the lemmatized verb and possible token-level or type-level classifications (e.g. whether it is considered a mental verb). With *types*3 we can then start to explore the use of the construction from different perspectives: we can see not only the number of types but also the number of hapax legomena, and we can compare the numbers with random subcorpora that have the same number of running words or the same number of BE *going to* V instances. These measures are similar to the productivity measures proposed by Baayen (1993, 2009) but designed for comparability across subcorpora of unequal size. To analyse the proportion of types of a specific kind (such as mental verbs) out of all types over time, we can also compare subcorpora with the same number of types. Finally, we can compare token frequencies of BE *going to* V in subcorpora with the same number of running words. We can explore different periodizations, zoom into various subcorpora, and visualize the results both over time and across genders.
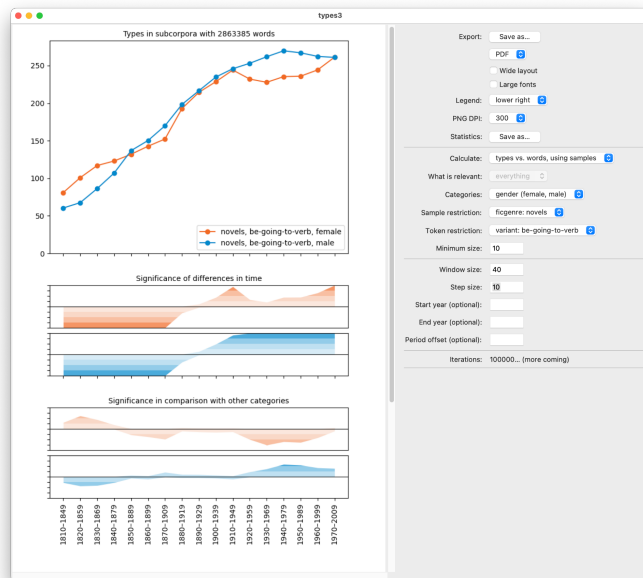
Figure 1. The user interface of *types*3 on macOS

In addition to the features discussed above, *types*3 differs from prior tools in terms of the user experience and the way the results are visualized and presented. First, the entire tool is developed for exploring variation in productivity over time, and all visualizations aim at providing a clear overview of diachronic trends and their statistical significance. Second, even though resampling statistics in general require a significant amount of computation (the results presented here are based on 1 million randomly constructed subcorpora), *types*3 makes it possible to conduct interactive exploration without doing any computation in advance: the user can adjust the settings, and approximate results based on a smaller number of random samples are presented almost instantly, while more accurate statistics are automatically computed in the background. The computationally intensive parts of the tool are written in Rust, while the user interface is written in Python; the program is freely available for download.[9]

4.1.4 *Overall trends*

Figure 2 shows the results that we obtain when we compare the type frequencies of the verb slot in the BE *going to* V construction in terms of the number of running words in the corpus. In the trend view on the left, *types*3 shows how the number of types evolves over time. The trend curve is computed as follows:

---

[9] https://github.com/suomela/types3

First, we find the period with the smallest number of words; in this case it was the period 1810–1849 with 18.8 million words. Then for each period we randomly reorder the texts, and see how many types we accumulate in the first 18.8 million words (and as the cutoff point is typically in the middle of a text, we record the number of types both before and after the final text). We repeat this process for each period 1 million times, and take the average number of types. This way we will get a trend curve that is useful for human inspection and easy to interpret: roughly speaking, it shows how many types we get if we take 18.8 million words of text from each period. This approach is similar to the one used by Berg (2021), and the idea of taking a sample of equal size from each subcorpus (though not the resampling and averaging) seems to have been first presented by Gaeta & Ricca (2006); see further Säily *et al.* (2024: 11).
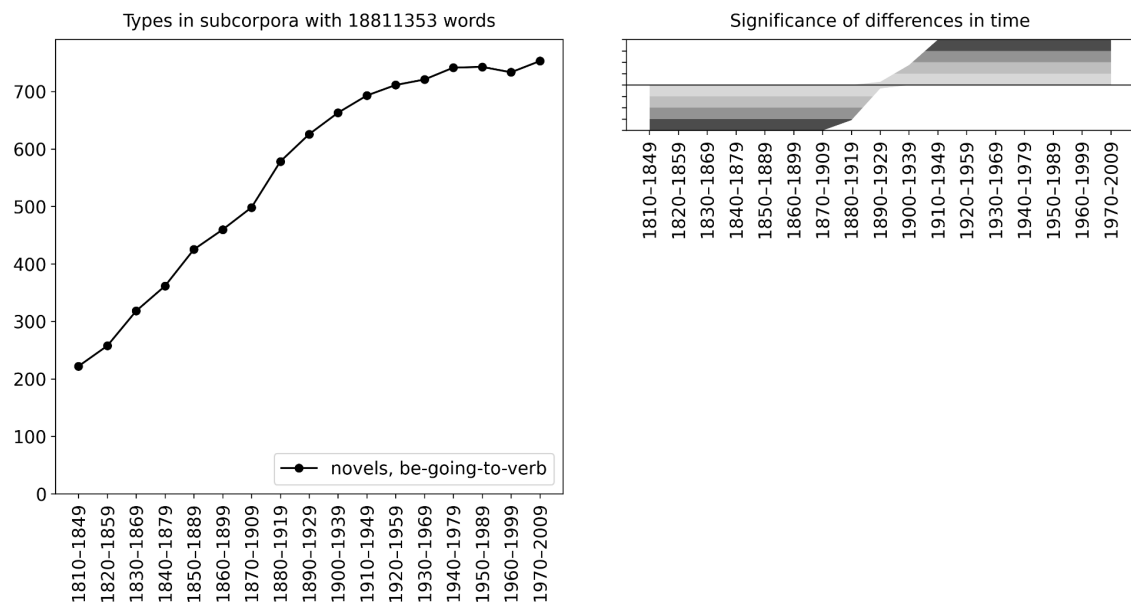

Figure 2. Change in the type frequency of BE *going to* V by the number of running words in COHA novels

However, while there is a clear increase in the number of types in the trend curve, it does not yet tell us whether the findings are statistically significant – we would never expect a perfectly flat line in any real-world corpus, but is the trend that we see here strong enough that we can say it is unlikely to be a random artefact unrelated to any genuine diachronic trend? This is where we need the significance curve that is shown on the right in figure 2. Here we have used permutation testing for each period to test the following hypotheses: (1) the number of types in this period is significantly high in

comparison with the corpus as a whole, and (2) the number of types in this period is significantly low in comparison with the corpus as a whole (the null hypothesis being that the time period of a text does not influence the number of types).

To do such testing for one period, for example, period 1940–1979, we take *all* data that we have in this period (it turns out to be 29.1 million words and 916 types); note that this is different from the way the trend curve is produced. We then randomly sample texts from the entire corpus so that we accumulate a total of 29.1 million words, and calculate how many types we get that way. We repeat this process 1 million times and see how many times we get fewer than 916 types this way, and how many times we get more than 916 types this way (again keeping in mind that the exact cutoff point may be in the middle of a text, and we need to correct in the right direction so as not to overestimate significance). It turns out that almost all 29.1-million-word collections of text from the entire corpus have fewer than 916 types; we can conclude that the type diversity in 1940–1979 is indeed significantly high in comparison with the corpus as a whole, with $p < 0.0001$. This is indicated in the significance plot with the dark shading for 1940–1979; different shades correspond to $p = 0.1, 0.01, 0.001$, and 0.0001. We can see from the significance plot that early periods (until around 1870–1909) have significantly few types in comparison with the corpus as a whole, while later periods (starting around 1910–1949) have significantly many types.[10] The basic principle of using permutation testing to assess significance this way in the context of type diversity has been used in prior work by Säily & Suomela (2017), but the way we visualize the results over time is, to our knowledge, new.

4.1.5 *Gender differences*

Figure 2 indicates that the productivity of the construction increases over time in the corpus as a whole. Next, let us consider the social factor of gender. Figure 3 shows that the productivity increases for both male and female authors. This is evident in both the trend view on the left and the significance view on the top right, which shows that the productivity is significantly low for men of the early periods compared to all men, and significantly high for men of the later periods compared to all men, and the same holds for women. On the other hand, the

---

[10] We note that we do multiple hypothesis testing here. If we aim at an overall confidence level of 0.05, and we test 2 hypotheses for each of the 17 periods, with a simple Bonferroni correction we would consider results with $p < 0.0015$ statistically significant. This approximately corresponds to the top of the third band in the significance plot.

trend view of figure 3 does not indicate a consistent gender difference over time. Rather, women appear to lead the change at first, after which both genders advance in tandem, with women falling behind by the latter half of the twentieth century. This is corroborated by the significance view on the bottom right, which shows that the productivity is significantly low for women of the later periods compared to men of the same periods.
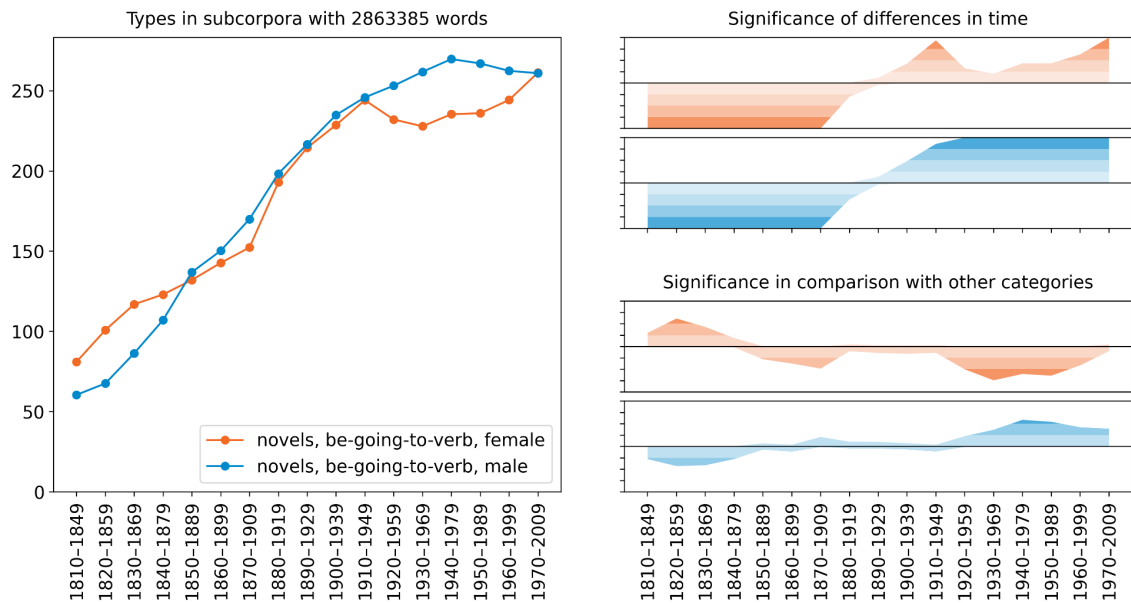

Figure 3. Gender variation and change in the type frequency of BE *going to* V by the number of running words in COHA novels

4.1.6 *Focus on diversity*

Figures 2 and 3 consider the number of types in terms of the number of running words in the corpus. As noted by Säily (2011: 148), this in essence conflates two measures: how often and how diversely the construction is used. If we want to focus our attention on diversity alone, we can assess the number of different verb types that we encounter in terms of the number of instances (tokens) of the construction: when the construction is used, how diversely is it used? This perspective is presented in figure 4. Interestingly, by this measure the productivity of the construction displays no increase whatsoever over time. In fact, the productivity becomes significantly low in the early twentieth century compared to the corpus as a whole.
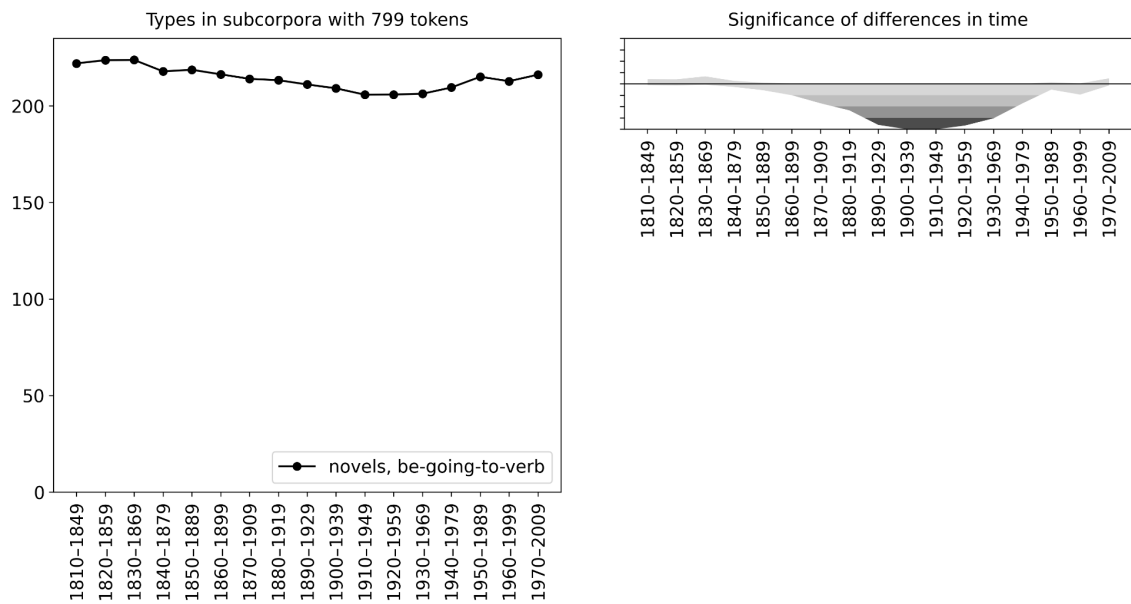
Figure 4. Change in the type frequency of BE *going to* V by the number of tokens of the construction in COHA novels

Figure 5 shows the situation by gender, and this time there is a consistent gender difference: men tend to use the construction more productively than women throughout the two centuries. While the trend view indicates that this difference decreases over time, there are still significant gender differences in periods representing the late twentieth century. This is probably because we have more data from these periods, which enables us to reject the null hypothesis with more certainty, even if the difference might not be as large. When we consider hapax legomena rather than types, the trend remains similar, indicating that male and female authors differ not only in their extent of use of different verbs in the construction but also in what Baayen (2009: 902) calls potential productivity, which estimates the construction's potential for expansion to new verbs.
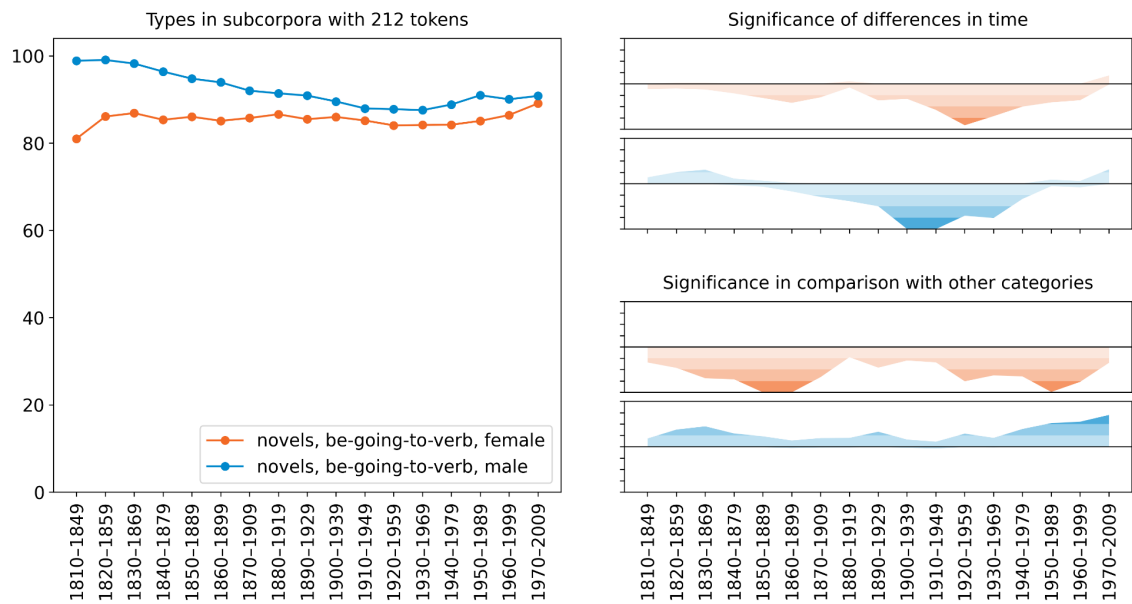
Figure 5. Gender variation and change in the type frequency of BE *going to* V
by the number of tokens of the construction in COHA novels

4.1.7 *Internal factors*

Let us next move on to the internal factors, which paint a rather different picture of variation and change in the productivity of the construction. Here we analyse the proportion of verb types representing the factor in question out of all verb types used in the construction. Firstly, the proportion of mental verb types out of all verb types in the construction shows an increasing trend over time, which is supported by the proportion being significantly low in the initial periods, particularly 1830–1869 (figure 6). Moreover, the change is clearly led by women, which is the opposite of what we saw in the overall type frequency. While the gender difference is not very significant, it is quite consistent, which speaks for its being a real phenomenon in the corpus.
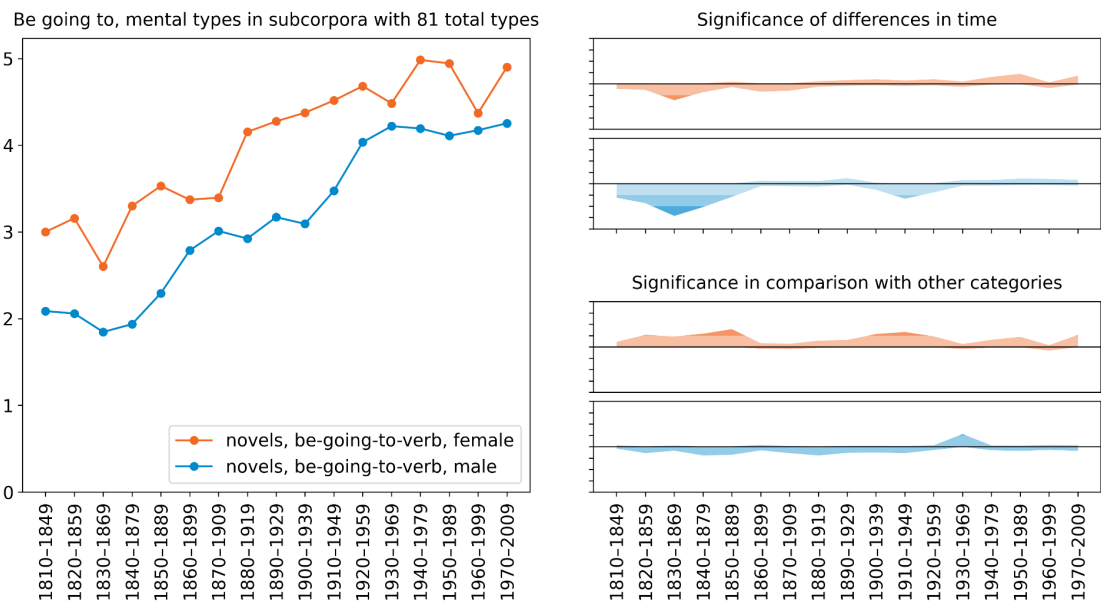
Figure 6. Gender variation and change in the proportion of mental verb types
in the BE *going to* V construction in COHA novels

The proportion of verb types used with the inanimate subject *it* behaves in a similar manner (figure 7). There is an increasing trend over time, and the process appears to be female-led, although this is somewhat less clear in the trend line than in the case of mental verbs. Both of these internal factors indicate increasing productivity of the 'prediction' sense of the construction, as in examples (4) and (5) in section 2 above.
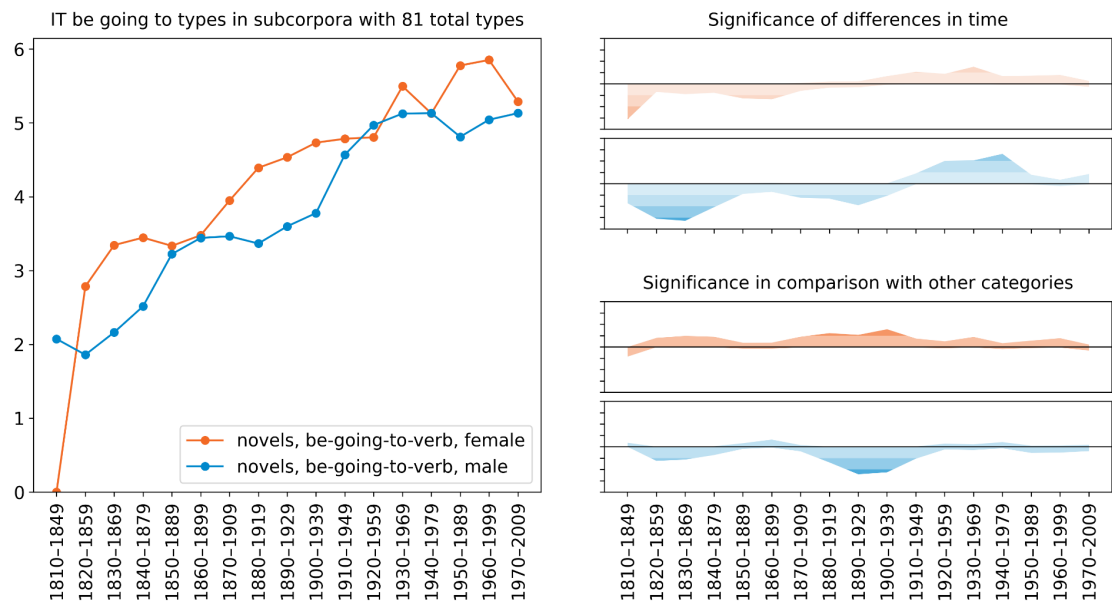
Figure 7. Gender variation and change in the proportion of verb types used with the inanimate subject *it* in the BE *going to* V construction in COHA novels

Figures 8 and 9 focus on the proportion of verb types used with passive voice (the *be*-passive) in the construction, as in example (6). The latter figure indicates that overall, the proportion increases over time, peaking around 1930–1969. Figure 8 displays no consistent gender difference, although women's usage is somewhat high in the late 19th to the early 20th centuries (1870–1909, 1880–1929 and 1890–1939). Furthermore, it is only men for whom we see an increasing trend over time, excepting the first few decades. Women seem to rather have a decreasing trend, but this is not supported by the significance analysis.
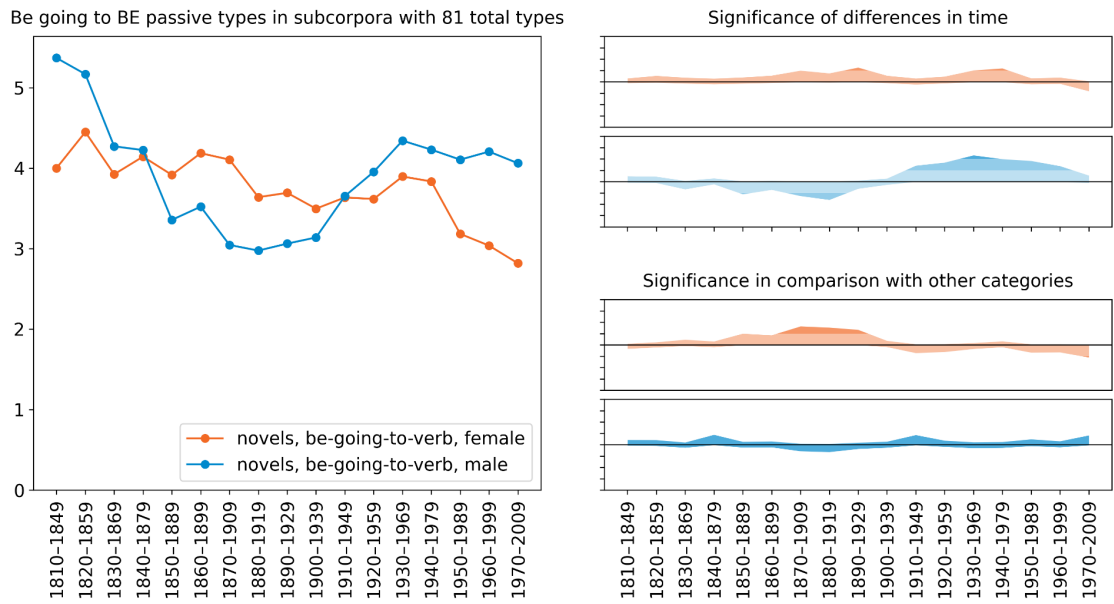
Figure 8. Gender variation and change in the proportion of verb types used
with passive voice in the BE *going to* V construction in COHA novels
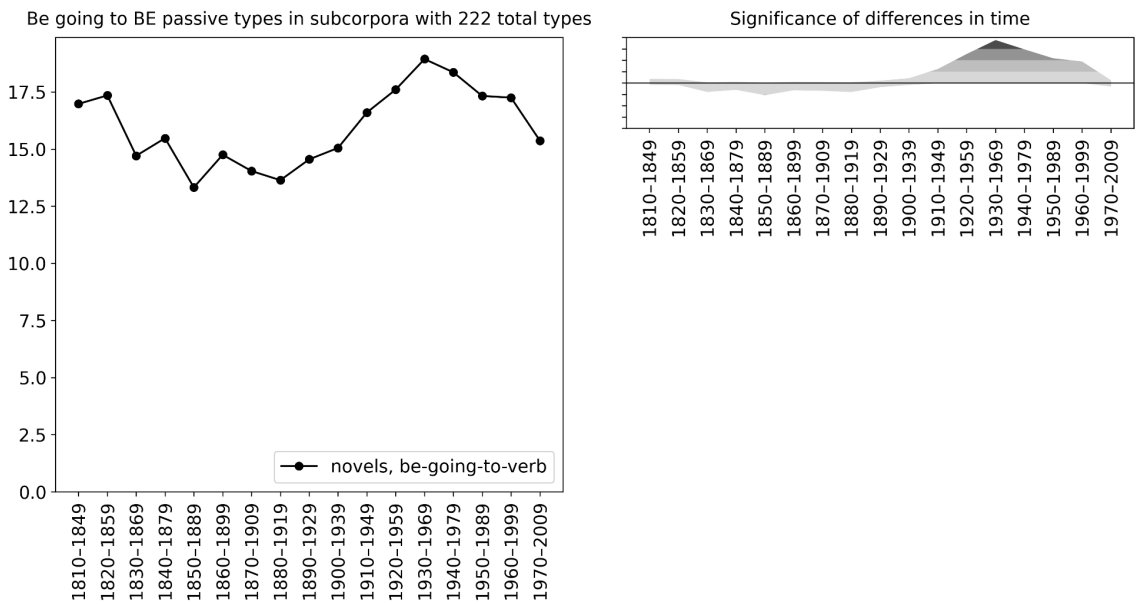

Figure 9. Change in the proportion of verb types used with passive voice in the
BE *going to* V construction in COHA novels

4.2 *Distributional semantic analysis*

Our analysis of the BE *going to* V construction so far has enabled us to identify quantitative differences between female and male writers in their use of the construction. In particular, we detected differences in the type frequency of the verb slot, which is traditionally considered to provide an indication of the degree of openness of a slot. However, a growing number of studies suggest that it is also important to consider what *kind* of types are in the distribution of a construction. For instance, Suttle & Goldberg (2011) argue that it is primarily variability, i.e. the diversity in the types attested with a construction, that drives the productivity of constructions. The underlying intuition is that some constructions might be restricted to only certain specific semantic areas and others might be compatible with a wide range of meanings, but these differences may not be reflected in type frequency counts.

Against this backdrop, in this section we explore the semantics of the items used in the verb slot of the BE *going to* V construction, and we examine changes to the semantic areas occupied by the construction. In particular, we seek to gauge to what extent the construction becomes increasingly used with verbs that do not align with the 'motion with purpose' source of the constructions, and whether this differs between men and women. In this, we expand on the study of mental verbs reported in the last section (see also Wu *et al.* 2016) by looking at a wider range of semantic areas, and we do so in a more data-driven way. To achieve this, we follow Perek (2016, 2018) and Hilpert & Perek (2022) in adopting a distributional approach to lexical semantics for the study of syntactic productivity. We use distributional semantic representations to measure the semantic spread of the construction over time and between genders.

4.2.1 *Distributional semantic model*

Distributional semantics is an approach to lexical semantics that aims to capture the meaning of words through their lexical collocates in large text corpora, following Firth's (1957: 11) oft-cited intuition that 'you shall know a word by the company it keeps'. It is based on the idea that semantically similar words are expected to have the same collocates. For instance, the verb *drink* and *sip* refer to similar actions (ingestion of fluids), and therefore are expected to co-occur with the same set of words, for instance words for beverages (*wine*, *water*, *coffee*, *beer*), containers (*cup*, *glass*, *bottle*), as well as words related to drinking or dining practices (*restaurant*, *bar*, *table*, *party*, etc.). In distributional semantics, words are considered similar in meaning to the extent that they are

similar in distribution. In a Distributional Semantic Model (hereafter DSM), the meaning of each word is typically captured as a vector, i.e. an array of numerical values, from which various kinds of quantitative information can be derived. In particular, the semantic similarity between words can be measured by quantifying the similarity between vectors; studies in the field typically achieve this by means of the cosine similarity measure, i.e. the cosine of the angle between two vectors, which thus captures to what extent two vectors 'point' in the same direction in the hyper-dimensional semantic space (cf. Turney & Pantel 2010). The main advantages of the cosine measure are its conceptual simplicity and low computational complexity, and it has been shown to achieve the best results for this purpose over other available measures (cf. Bullinaria & Levy 2007).

For this study, we used a DSM from Perek (2022) created with word2vec, a computational approach to distributional semantics that aims to quantify the relation between a word and its contexts of occurrence in a corpus by means of a neural network (Mikolov *et al.* 2013); the semantic vectors are derived from the weights of nodes in the layers of the neural network. This model was trained on a PoS-tagged version of the whole COHA corpus; as such, it provides a kind of 'historical average' of the meaning of words over the time span of the corpus. While we would ideally need to use different models trained on different sections of the corpus to account for possible lexical semantic change, using a single model means that we use the maximum amount of data for each word, leading to more robust and reliable semantic representations, and it also presents the advantage of facilitating comparisons between periods, since the location of words with respect to each other in the semantic space is held constant. New meanings of verbs have obviously appeared over the past 200 years, but given the relatively shallow time depth and the advanced stage of standardization in English over the period of interest, we can be reasonably confident that the overall meaning of most verbs in our sample has not changed dramatically and that the 'historical average' approach should not substantially distort our findings.

The DSM was created with the SkipGram algorithm of word2vec using 1000 dimensions. Given a word, SkipGram attempts to predict its context of occurrence, defined here as the surrounding words in a 2-word window (i.e. 2 words to the left and 2 words to the right); words are considered similar to the extent that they predict similar contexts. Using the lemma annotations of the corpus, all forms of verbs were grouped together as one word in the model; in other words, each verb was assigned a single vector corresponding to the distribution of all forms of the verb taken together.

4.2.2 *Measuring the semantic range of constructions*

The method pioneered by Perek (2016, 2018) consists in extracting from a DSM pairwise cosine similarity scores between verbs occurring in a construction at different points in time, and using this information to plot these verbs in two dimensions with multidimensional scaling or *t*-SNE (Van der Maaten & Hinton 2008). These distributional semantic plots are used to visualize the semantic space populated by these verbs and identify semantically coherent classes of verbs in different time periods, thus showing how the semantic spread of a construction changes over time. However, in its original form, this way of assessing the productivity of constructions suffers from the same issues as measuring type frequency, as discussed in section 4.1.1: since the number of types does not vary linearly with sample size, we cannot compare semantic plots of types found in samples of different sizes, as we cannot be sure that the presence or absence of certain types can genuinely be interpreted as differences in productivity or is simply due to differences in sample size. A possible solution to this issue is to randomly select tokens of the construction so as to match sample size for every time period. For instance, Perek (2018) normalizes samples of the *way*-construction in the entire COHA corpus by randomly selecting texts from each decade in the corpus so that the set of texts for each decade matches the size of the smallest decade in the entire corpus. The tokens of the construction and thus the corresponding types to be used for the distributional semantic plots are then extracted from the selected texts. However, this approach is only appropriate when the corpus is not too imbalanced, i.e. there is relatively little variation between samples, and in particular the difference in size between the smallest sample and the other samples is not too great. Unfortunately, neither of these conditions are met by the gender-annotated COHA corpus of novels we are using in this study, as described in section 3. Hence, a different approach is needed.

In this study, we adapted the general logic of *types*3 to the study of the semantic range of a construction in diachrony. In the previous work cited above, semantic change in a construction was assessed by considering individual types and how they relate to the rest of the distribution, but such a qualitative assessment is not possible over randomly selected samples, as specific types may vary widely from one sample to the next, and the range of individual types cannot be straightforwardly averaged over samples in the same way that overall type frequency is in types3. The solution to this issue that we describe below is to sort the types into discrete semantic categories, and then to calculate average type counts in each category across random samples. The distribution of the average type counts over semantic categories provides a representation of the

average 'semantic spread' of the construction, which can be used to quantify semantic variation in the construction, both between time periods and between genders (see Hilpert & Perek 2022 for a similar approach).

This approach was implemented as follows; all the steps in this procedure were done in R. We used the same dataset as in the *types*3 analysis presented above. From this data, we extracted randomly selected samples containing the same number of tokens of the construction. We repeated this random sampling 1,000 times for each time period and each gender. To maximize the number of tokens in each sample, we use 40-year periods from 1820 to 1979, hence 1820–1859, 1860–1899, 1900–1939, and 1940–1979; changes happening in the three decades after 1979 are thus ignored. The number of tokens in each random sample is set to match the minimum sample size found in our dataset across the whole periodization, i.e. 534 (corresponding to the number of tokens contributed by female writers in the 1820–1859 period). As in *types*3, to preserve the textual units of the corpus, we sample tokens from the same text together.

We collected all the types attested in these random samples. We then extracted pairwise semantic similarity scores (using the cosine measure) between these 1419 types from the DSM described above, and we used these scores to automatically group types into semantic categories using cluster analysis, using the Partitioning Around Medoids (PAM) algorithm from the 'cluster' R package. This allowed us to identify semantic classes automatically in a data-driven way, without assuming any prior categorization.

As a variant of *k*-means, PAM requires the user to specify a target number of clusters when running the algorithm. The 'right' number of clusters is typically chosen by comparing different clustering solutions with different numbers of clusters, and deciding which one is most appropriate; our criteria for appropriateness were mostly qualitative.[11] For our present purposes, we need

---

[11] As is mentioned by many introductions to cluster analysis, there are also quantitative ways to determine the 'ideal' number of clusters, which generally aim to find the number that maximizes both the internal consistency of clusters and the variation between clusters, as measured by some mathematical criterion, for instance average silhouette width (cf. Levshina 2015: 311). However, this method did not lead to a satisfactory outcome in our case: as we increased the number of clusters, the increase in average silhouette width only started to slow down at 284 clusters, which is much too high a number to be useful to us in practice, as it results in very fine-grained semantic classes containing few items (between 1 and 17, with a mean of 5), making it less likely to detect differences in type frequency over time in such small classes, and there are too many clusters to allow for a manual inspection of each. Therefore, we chose to use a more qualitative approach by semi-arbitrarily setting the number of clusters to obtain a clustering solution that is appropriate for our purposes, following criteria described after this

to divide the verbs into clusters that are large enough for us to measure changes in type frequency, and at the same time are still semantically coherent, i.e. we are able to assign a relatively precise semantic definition that covers most members of each cluster. The total number of clusters should also not be too high, so that we are able to examine change in each cluster individually and compare it to other clusters. We tried clustering solutions of increasing numbers of clusters (namely, 10, 20, 30, and 40) and checked each solution qualitatively for these criteria. We found a 30-cluster solution to strike a good balance between a reasonably low number of clusters that at the same time were still largely semantically coherent. This number was also low enough to enable individual analysis. The full list of clusters with all their verbs can be found in the appendix.

For each of the 30 clusters of verbs thus identified, we calculated the average number of types attested in each period and for each gender across the 1000 samples. For instance, Sample 1 contains two types belonging to Cluster 1 in Period 1 (1820–1859) for men, Sample 2 two types as well, Sample 3 four types, etc., and when we average over all the samples, we find a mean of 2.824 types in Period 1 for men. The outcome of this procedure is two matrixes of average type counts per time period and per cluster, one for men and one for women. Because the total average type counts vary slightly from one period to the next, as a final step we converted the average type counts into type proportions, by dividing the type count of each cluster by the total type count of the relevant time period. This provides an indication of the prominence of each semantic class within the distribution of the construction. Two types of analysis can be applied to this data: (i) a general quantitative analysis comparing the distribution of the construction between periods and between genders, i.e. the columns of the matrixes, and (ii) a fine-grained, more qualitative analysis focusing on particular semantic classes and measuring type frequency variation in these classes for each gender. We present each type of analysis in turn.

4.2.3 *Overall results*

Table 2 and table 3 report similarity scores between the type distributions of the construction for different periods or genders as measured by Pearson's correlation coefficient (ρ). Pearson's ρ measures correlations between vectors and varies between -1 and 1, with -1 indicating a negative correlation, 1 a

---

footnote. As will be clear by the end of this section, this approach allowed us to successfully identify changes in type frequency that we can semantically interpret with respect to the meaning of the construction and its grammaticalization history.

positive correlation, and 0 no correlation; hence, the closer to 1 ρ is, the more similar the two distributions are. Table 2 reports comparisons between adjacent periods, for men and for women; in this table, a lower ρ indicates change over time, and a higher ρ indicates stability. Table 3 reports comparisons between genders in the same period; in this case, a lower ρ indicates divergence, and a higher ρ indicates similarity in usage.

Table 2. *Change in the semantic distribution of BE* going to *V for each gender, as measured by Pearson's ρ*

	Period 1 vs. Period 2	Period 2 vs. Period 3	Period 3 vs. Period 4
Men	0.937	0.975	0.982
Women	0.942	0.950	0.978

Table 3. *Comparison of the semantic distribution of BE* going to *V between genders over time, as measured by Pearson's ρ*

1820–1859	1860–1899	1900–1939	1940–1979
0.901	0.938	0.970	0.978

Table 2 shows that change is happening in the distribution of the construction from Period 1 to Period 2 to a similar extent for both genders, although the degree of change is arguably very modest, as ρ is over 0.9. This lines up with the finding that, by the early 19th century, the BE *going to* V construction has fully acquired its modern meaning (though not its full Present-Day English distribution) and is thus already open to a very wide range of verbs, leaving few semantic areas that have not yet been covered. At the same time, the amount of change decreases over time (i.e. ρ increases), especially for men, as distributions become more similar from one period to the next. Table 3 shows that the way men and women use the BE *going to* V construction is already very similar from the first period on, but it becomes even more similar throughout the 19th and early 20th century, which indicates that the two genders are converging in their use of the construction.

Figure 10 focuses on each of the 30 semantic classes individually, as it plots the variation in type count proportions in each class over time separately, for men (blue) and women (red). This allows us to identify which classes show gender differences in productivity, i.e. cases in which one gender uses more types in one class than the other gender at a given point in time.
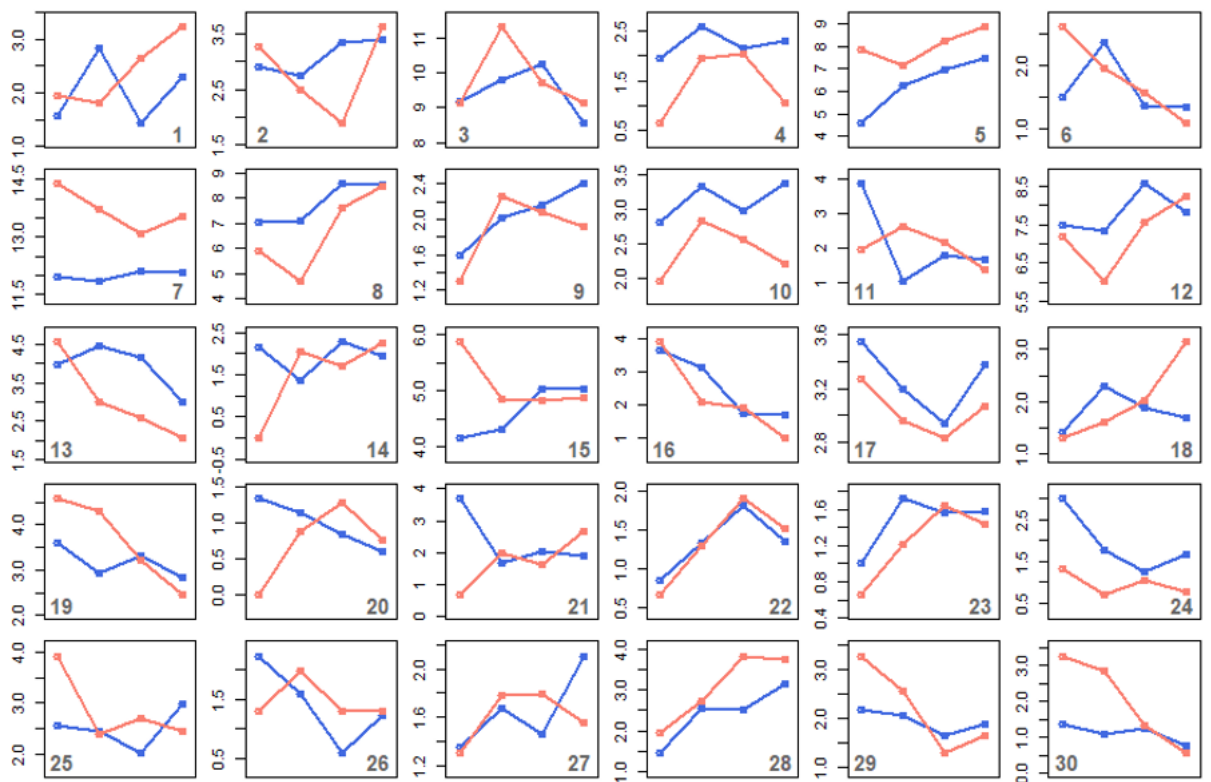
Figure 10. Variation in type proportions in each class over time, for men (blue) vs. women (red). Samples are taken from four periods: 1820–1859, 1860–1899, 1900–1939, 1940–1979

As can be seen in figure 10, while in some classes the type proportions are roughly similar between genders and follow the same trends (e.g. clusters 9, 16, 22), many others do show some gender differences. It is beyond the scope of this case study to discuss each of them in detail, and another reason why it might be difficult to do so anyway is that, while many of the cluster can indeed be interpreted semantically and can clearly be assigned to a particular semantic class of verbs, not all of them can; some correspond to a collection of various classes rather than a single coherent semantic category. Such an outcome is not unexpected given the low number of clusters and the high number of verbs, but it makes it difficult to interpret changes in type frequency for such heterogeneous classes. Therefore, we will only focus on a small selection of classes that do receive a clear semantic interpretation, but this should not be taken as an exhaustive representation of the changes happening to the distribution of the construction.

4.2.4 *Analysis of individual semantic classes*

Table 4 lists the six clusters chosen for the analysis, which we discuss in turn below. In each cluster, most verbs can be seen to correspond to a broad semantic class, identified in the first column, although there are also some outliers, listed in brackets below the main group. In some cases, even the outliers can be seen to relate somehow to the meaning of the group, for instance many of the outliers to the sound emission verbs (cluster 25) do refer to events that typically involve some sound, e.g. *clap*.

Table 4. *Semantic clusters of verbs chosen for the analysis*

Cluster 3: motion verbs	amble approach circle cross descend drive enter escort follow head herd hike hurry lead leave march meander motor move paddle parade pass ramble ride row run sail saunter shepherd shuffle skate sneak sprint stalk step stride stroll stumble swim travel tread trip trot trudge usher wade walk wander (+ breathe carry dance face listen live play seat sit stand stay wait watch)
Cluster 7: communication	advise ask cable call caution confide convince contact greet inform instruct interview invite lecture name notify phone recite remind retell say speak talk thank teach tease telephone tell warn (+ come find film fuck give hear know join look make meet read relate see show spot take think write)
Cluster 8: caused motion, change of state, manipulation	attach back bend break bring catch close crank cut dig drag draw grab grip hand haul hitch hold hook jab jerk jolt kick knot lift nudge open park pat pluck pick pin pinch place poke press pry pull push put ram rip scoop scrape set slit shake shoot shove slash snap split squeeze straighten strain strip suck swipe switch tap tense tie tow tuck turn twist unfasten untie unwind weave whittle withdraw (+ coax jump knit lean peek plop point reach square)
Cluster 25: sound emission	bubble cheer croak cough curse cry erupt holler faint howl jabber laugh mew plead pound rage roar scream shout shriek sing sneeze sound squawk swoon thunder weep whisper whistle yell (+ clap explode hug ring storm vomit)
Cluster 27: mental verbs (emotion type)	affect alarm amuse astonish attract bother confuse cloud depress disappoint disturb excite frighten hurt impress influence miss offend please rouse scare shock stir surprise trouble unhinge upset worry (+ accompany dictate interfere interrupt intrude)

Cluster 28: mental verbs (cognitive type)	accept acquaint appreciate believe describe discover excuse explain fear feel forget guess heed ignore interpret judge learn mention mind notice realize recognize regret remember respect satisfy solve suppose suspect understand wonder (+ belong exist materialize penetrate stop suit unravel)

First, cluster 3 contains verbs of motion, including many verbs encoding manner of motion, e.g. *amble*, *meander*, *sneak*, *run*, as exemplified by (7) and (8). We would argue that such a group is especially relevant to the study of the productivity of BE *going to* V, given the semantic source of the construction. In the original 'motion with purpose' meaning of BE *going to* V, it would seem incongruent to make the purpose of motion another form of motion: one does not normally go somewhere in order to go somewhere else. Hence, occurrence of verbs of motion in the construction can be seen as a sign that it has grammaticalized into its futurity meaning (cf. Traugott & Trousdale 2013: 118), and increased productivity in this class indicates a weakening of lexical persistence effects and further abstraction of the constructional meaning. As seen in figure 10, women seem to be ahead of men in using this class productively in the middle of the 19th century, though men quickly catch up in the 20th century. If we examine this cluster of verbs in *types*3 by looking at the proportion of cluster 3 types out of all verb types in the construction over time, we obtain very similar results. This can be taken to suggest that the construction initially shows a higher degree of grammaticalization for female writers than for male writers, at least with respect to this class.

(7) At present I **am going to meander** through the lancers with Mrs. Featherbrain. (COHA, 1844)

(8) I believe he **is going to walk** to the station; he certainly is crazy. (COHA, 1886)

Similar comments can be made in relation to cluster 27 and cluster 28. Both clusters correspond to a general class of mental verbs, and indeed many verbs from Halliday & Matthiessen's (2014: 256–7) list are found in them. The two clusters can be distinguished in that the former mostly corresponds to what Halliday & Matthiessen term the emotion type (e.g. *fear*, *impress*, *shock*), and the latter mostly corresponds to the cognitive type (e.g. *believe*, *suspect*, *understand*), as exemplified by (9) and (10) respectively. As argued above, these verbs too are relevant to the grammaticalization of BE *going to* V, as they are predicates that are a priori not tied to spatial location, and thus the 'motion

with purpose' source meaning of the construction is not readily compatible with them. Women seem to be slightly more productive with these verbs than men, albeit in different ways. In the emotion type, women are ahead in Periods 2 and 3 (i.e. from 1860 to 1939), but they are overtaken by men in the mid 20th century. In the cognitive type, the female preference is initially quite small, but women are consistently in the lead, and the gap between men and women markedly widens after 1940, with women using the cognitive type much more productively. These findings roughly follow the same trends as the analysis using the *types*3 software reported in section 4.1.7, but they also suggest that not all mental verbs behave exactly in the same way, which shows the benefits of the distributional semantic approach in identifying verb classes. It is not clear how the difference between these two groups should be explained, but it could be related to the fact that women also seem to be ahead of men in using the construction with an inanimate subject, as reported earlier (cf. figure 7): most verbs in the emotion group take a stimulus argument as subject that can be inanimate, and thus are more likely to have an inanimate subject than the verbs in the cognitive group, which predominantly take an animate experiencer as subject.

(9)   He began to see how this misfortune **was going to affect** him. (COHA, 1895)

(10)   If you're a good boy maybe **we're going to forget** we had this little misunderstanding. (COHA, 1920)

Cluster 7 and cluster 25 exemplify two related domains: the former contains communication verbs (e.g. *ask*, *talk*, *warn*), the latter includes some manner of speaking verbs (e.g. *whisper*, *yell*) as well as various verbs conveying sound emission, which can often be used in a communication or social interaction sense (e.g. *croak*, *laugh*, *roar*, *scream*); see examples (11) through (13). Interpreting these verbs with respect to the degree of grammaticalization of the construction is less straightforward than with the classes discussed so far, but since many of these verbs are of a relatively abstract type and their meaning is not necessarily tied to spatial location, it is fair to assume that they at least occupy an intermediate position between verbs that lend themselves well to the 'motion with purpose' interpretation, and verbs that are incompatible with it. Although the general diachronic trend in both classes is one of decline relative to other classes, women seem to use both classes markedly more productively than men, at all times for verbs of communication, and only up until the 20th century for verbs of sound emission.

(11) "Pomp, sit down," said Frank. "I **am going to talk** to you seriously." (COHA, 1864)

(12) "Keep still, Mortimer," she said, "**I'm going to whisper** right in your ear." (COHA, 1911)

(13) I will not tell you if you **are going to laugh**. (COHA, 1855)

Women seem to show a higher degree of productivity in the classes discussed so far, but this is not always the case, and in some clusters it is men who are in the lead. A particularly good example of this is cluster 8, a diverse class of loosely related items which are unified by a high degree of physicality: verbs of caused motion (e.g. *bring*, *place*, *pull*), change of state verbs (e.g. *break*, *cut*, *open*), and verbs of forceful contact and manipulation (e.g. *attach*, *kick*, *squeeze*). Men consistently use this class more productively than women up until the mid 20th century. With regards to the grammaticalization of the construction, these verbs arguably lend themselves more readily to the original 'motion with purpose' meaning than the other classes discussed so far, as they are simply more tied to the physical realm than any of the other classes discussed so far. Hence, the productivity of the construction in this class can be interpreted as an effect of lexical persistence, and conversely, it does not suggest a higher degree of grammaticalization of the construction. A possible explanation of the gender difference in this class could thus be that men generally use the construction more conservatively than women, which directly mirrors the more innovative behaviour displayed by women in the classes discussed above.

(14) It was rumored that the Fenians here in Milwaukee **were going to place** dynamite bombs at every house where an Englishman lived. (COHA, 1883)

(15) He can't **be going to cut** willows. (COHA, 1884)

(16) I **am going to catch** fish, and sell them in town, sir. (COHA, 1860)

5 DISCUSSION AND CONCLUSION

Our findings indicate that contrary to the results of Wu *et al.* (2016) on token frequency, the overall type frequency of BE *going to* V does not increase over time in the Late Modern and Present-Day American English of our dataset, and there is even a slight decrease in the early twentieth century. It would then seem that productivity-wise, the grammaticalization process has stalled since the

nineteenth century. Moreover, we have found that men's usage of the construction is more productive than women's, with a converging trend over time.

However, the situation is reversed when we consider intralinguistic factors related to subjectification: the proportion of mental verb types, verb types used with the inanimate subject *it*, and verb types used with passive voice in the construction all exhibit an increasing trend over time. Moreover, the first two changes appear to be led by women. Thus, by zooming in on the internal factors, we can see that the ongoing process of subjectification manifests itself not only as increases in token frequency (Wu *et al.* 2016; Budts & Petré 2016) but also in productivity, and far from stagnating, the grammaticalization process is still ongoing from this perspective. The fact that two out of three of the proportional changes are led by women is in accordance with previous sociolinguistic research (e.g. Labov 2001), and the contrary findings from morphological productivity (e.g. Säily 2014; Säily *et al.* 2024) are not replicated for syntactic productivity. More research is however needed to verify these tendencies regarding different levels of linguistic organization.

Supporting and augmenting the results on mental verb types, our data-driven type-based semantic analysis also identifies areas of growth. These include a cluster involving verbs of motion, which indicates grammaticalization in terms of futurity, as well as two clusters of mental verbs, which roughly map onto Halliday & Matthiessen's (2014: 256–257) emotive and cognitive categories. The emotion verbs in particular could be linked to the 'prediction' sense indicating subjectification. All of these changes appear to be female-led, although male authors catch up in the twentieth century first in terms of motion and then emotion.

If we assume that the fiction data used in this study is a faithful reflection of men and women's knowledge and usage of the construction, in a DCxG account our results would suggest slightly different representations of the construction over time for these two groups of speakers. Both the higher prominence of inanimate subjects and the higher diversity of verbs in particular semantic classes (namely mental verbs as well as other narrow classes identified in the distributional semantic analysis) could indicate that the construction is initially slightly more schematic for female authors than male authors (Perek 2020), in that it more readily sanctions uses that are in line with the more general 'prediction' sense, but not with the more narrow 'intention sense', much less with the original 'motion with purpose' meaning of BE *going to* V. More research is certainly needed to confirm this interpretation, for instance by including a wider range of internal factors.

We have seen that at the stage of grammaticalization of BE *going to* V displayed by our dataset, the overall type diversity of the construction stagnates but internal factors linked to grammaticalization indicate increasing productivity. This implies that internal factors are important to account for in analyses of productivity. Furthermore, we have found consistent gender differences in our dataset. While many of the changes seem to be female-led, there is a caveat here: male and female authors may tend to write in different subgenres of novels, which could influence the results. For instance, if women wrote more in genres characterized by a personally involved style, they would have more occasion to utilize mental verbs in general. When comparing the overall token frequencies of these verbs by gender, however, we observed no consistent correlation with their usage within the construction. Therefore, gender cannot be ignored as a possible factor, and many diachronic constructional analyses would benefit from considering it (cf. Säily *et al.* 2024).

Author's address:
Tanja Säily
Department of Languages, Faculty of Arts
P.O. Box 24
00014 University of Helsinki
Finland
E-mail: tanja.saily@helsinki.fi

Florent Perek
Department of English Language and Linguistics
University of Birmingham, Edgbaston
Birmingham B15 2TT
United Kingdom
E-mail: f.b.perek@bham.ac.uk

Jukka Suomela
Department of Computer Science, Aalto University
P.O. Box 15400
00076 Aalto
Finland
E-mail: jukka.suomela@aalto.fi

REFERENCES

Baayen, R. H. 1993. On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer.

Baayen, R. H. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 899–919. Berlin: Mouton de Gruyter.

Berg, Kristian. 2021. Productivity, vocabulary size, and new words. A response to Säily (2016). *Corpus Linguistics and Linguistic Theory* 17(1), 177–87.

Budts, Sara & Peter Petré. 2016. Reading the intentions of *be going to*: On the subjectification of future markers. *Folia Linguistica Historica* 37, 1–32.

Bullinaria, John A. & Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* 39(3), 510–26.

Bybee, Joan L. & William Pagliuca. 1987. The evolution of future meaning. In Anna Giacalone-Ramat, Onofrio Carruba & Giuliano Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics* (Current Issues in Linguistic Theory 48), 109–22. Amsterdam: John Benjamins.

Bybee, Joan L., Willian Pagliuca & Revere D. Perkins. 1991. Back to the future. In Bernd Heine & Elizabeth Closs Traugott (eds.), *Approaches to grammaticalization, volume II: Focus on types of grammatical markers* (Typological Studies in Language 19:2), 17–58. Amsterdam: John Benjamins.

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Danchev, Andrei & Merja Kytö. 1994. The construction *be going to* + infinitive in Early Modern English. In Dieter Kastovsky (ed.), *Studies in Early Modern English* (Topics in English Linguistics), 59–77. Berlin: Mouton de Gruyter.

Davies, Mark. 2010–. *The Corpus of Historical American English*: 400 million words, 1810–2009. https://www.english-corpora.org/coha/

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In *Studies in linguistic analysis (Special volume of the Philological Society)*, 1–32. Oxford: Blackwell.

Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1), 57–89.

Halliday, M. A. K. & Christian M. I. M. Matthiessen (2014). *Halliday's introduction to functional grammar*, 4th edition. Routledge.

Hilpert, Martin. 2008. *Germanic future constructions: A usage-based approach to language change*. Amsterdam: John Benjamins.

Hilpert, Martin & Florent Perek. 2022. You don't get to see that every day: On the development of permissive *get*. *Constructions and Frames* 14(1), 14–41.

Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization? A look from its fringes and its components* (Trends in Linguistics: Studies and Monographs 158), 21–44. Berlin: De Gruyter Mouton.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.

Labov, William. 1994. *Principles of linguistic change, volume 1: Internal factors* (Language in Society 20). Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change, volume 2: Social factors* (Language in Society 29). Malden, MA: Blackwell.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv: 1301.3781.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson Education.

Nuñez Pertejo, Paloma. 1999. 'Be going to' + infinitive: Origin and development. Some relevant cases from the Helsinki Corpus. *Studia Neophilologica* 71(2), 135–42.

Öhman, Emily, Tanja Säily & Mikko Laitinen. 2019. Towards the inevitable demise of *everybody*? A multifactorial analysis of *-one/-body/-man* variation in indefinite pronouns in historical American English. 40th Annual Conference of the International Computer Archive of Modern and Medieval English (ICAME 40), Neuchâtel, Switzerland, June 2019. https://tanjasaily.fi/talks/icame40_ohman_et_al_2019.pdf

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1), 149–88.

Perek, Florent. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1), 65–97.

Perek, Florent. 2020. Productivity and schematicity in constructional change. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and networks in Diachronic Construction Grammar*, 141–166. Amsterdam: John Benjamins.

Perek, Florent. 2022. Distributional Semantic Models for English verbs and nouns. OSF. doi:10.17605/OSF.IO/N324F. osf.io/n324f.

Petré, Peter & Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4), 867–901.

Rodríguez-Puente, Paula, Tanja Säily & Jukka Suomela. 2022. New methods for analysing diachronic suffix competition across registers: How *-ity* gained ground on *-ness* in Early Modern English. *International Journal of Corpus Linguistics* 27(4), 506–28. Special issue, *Corpus studies of language through time*, ed. by Tony McEnery, Gavin Brookes & Isobelle Clarke.

Säily, Tanja. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1), 119–41.

Säily, Tanja. 2014. *Sociolinguistic variation in English derivational productivity: Studies and methods in diachronic corpus linguistics* (Mémoires de la Société Néophilologique de Helsinki XCIV). Helsinki: Société Néophilologique.

Säily, Tanja, Martin Hilpert & Jukka Suomela. 2024. New approaches to investigating change in derivational productivity: Gender and internal factors in the development of *-ity* and *-ness*, 1600–1800. In Sarah Buschfeld, Patricia Ronan, Theresa Neumaier, Andreas Weilinghoff & Lisa Westermayer (eds.), *Crossing boundaries through corpora: Innovative approaches to corpus linguistics* (Studies in Corpus Linguistics 119), 8–40. Amsterdam: John Benjamins.

Säily, Tanja & Jukka Suomela. 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. In Antoinette Renouf & Andrew Kehoe (eds.), *Corpus linguistics: Refinements and reassessments* (Language and Computers: Studies in Practical Linguistics 69), 87–109. Amsterdam: Rodopi.

Säily, Tanja & Jukka Suomela. 2017. *types*2: Exploring word-frequency differences in corpora. In Turo Hiltunen, Joe McVeigh & Tanja Säily (eds.), *Big and rich data in English corpus linguistics: Methods and explorations* (Studies in Variation, Contacts and Change in English 19). Helsinki: VARIENG. https://urn.fi/URN:NBN:fi:varieng:series-19-10

Säily, Tanja & Turo Vartiainen. Forthcoming. Historical linguistics. In Michaela Mahlberg & Gavin Brookes (eds.), *Bloomsbury handbook of corpus linguistics*. Bloomsbury.

Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6), 1237–69.

Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds.), *Subjectification, intersubjectification and grammaticalization* (Topics in English Linguistics 66), 29–71. Berlin: De Gruyter Mouton.

Traugott, Elizabeth Closs. 2012. On the persistence of ambiguous linguistic context over time: Implications for corpus research on micro-changes. In Joybrato Mukherjee & Magnus Huber (eds.), *Corpus linguistics and variation in English: Theory and description*, 231–46. Amsterdam: Rodopi.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Turney, Peter & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–88.

Van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 2579–605.

Wu, Junhui, Qingshun He & Guangwu Feng. 2016. Rethinking the grammaticalization of future *be going to*: A corpus-based approach. *Journal of Quantitative Linguistics* 23(4), 317–41.

APPENDIX: SEMANTIC CLUSTERS OF VERBS USED IN THE DISTRIBUTIONAL SEMANTIC ANALYSIS

1	accost adore allude apologize baptize befall befriend beggar behead besiege betroth birth blindfold blush born botch bribe broach christen congratulate convict dazzle delight desert disinherit divorce dog elope exhaust fascinate flirt foreclose graduate haunt horrify indict induct jilt kiss marry miscarry nab nominate overtake overwhelm perjure quiz rescue sass serenade shame sober stab tempt trick waylay wed
2	abandon abolish annihilate attack blight bomb capture confiscate convert crumble crush deprive destroy devour dismantle displace disrupt engulf erase expose exterminate extinguish gut harm intercept invade jeopardize jettison kill lose outgrow outlive overrun overturn perish poison preserve raid raze rebuild render replace restore revolutionize rout ruin sacrifice seize smash spoil strangle survive swamp undo wither wreck
3	amble approach breathe carry circle cross dance descend drive enter escort face follow head herd hike hurry lead leave listen live march meander motor move paddle parade pass play ramble ride row run sail saunter seat shepherd shuffle sit skate sneak sprint stalk stand stay step stride stroll stumble swim travel tread trip trot trudge usher wade wait walk wander watch
4	avoid ban block cause cease control cure delay discourage dislodge eliminate encourage escape forbid force free guard keep occur omit permit prevent promote protect refrain relieve remedy remove resist save separate squelch stifle
5	beg begin can care choose dare decide deserve disagree dislike do enjoy expect get go hate help hope let like love mean need plan pretend promise refuse relish seek seem start tend try want will wish
6	appraise compare consider consult correct delve discuss examine explore identify insert inspect investigate observe question review scrutinize search sift study subdivide survey test view visit weigh
7	advise ask cable call caution come confide contact convince film find fuck give greet hear inform instruct interview invite join know lecture look make meet name notify phone read recite relate remind retold say see show speak spot take talk teach tease telephone tell thank think warn write
8	attach back bend break bring catch close coax crank cut dig drag draw grab grip hand haul hitch hold hook jab jerk jolt jump kick knit knot lean lift nudge open park pat peek pick pin pinch place plop pluck point poke press pry pull push put ram reach rip scoop scrape set shake shoot shove slash slit snap split square squeeze straighten strain strip suck swipe

	switch tap tense tie tow tuck turn twist unfasten untie unwind weave whittle withdraw
9	burst cast chuck clutch dash dump fetch flap fling flop fold heave hoist hurl leap level lock pelt pitch pounce pour rush scatter shower slap sling snatch spit spread spring stream stretch swallow sweep tear throw thrust toss utter whirl wrap
10	bounce buckle bump clamber climb collapse crash crawl crumple dip disappear dive dodge drift drop duck ease fall float fly fumble glide hop jot lapse lay loop maneuver melt nestle peep plunge pop race rattle retreat roll sag shift sink skim skip slide slip snuggle soar spill stoop swing tilt tip topple trail tumble tunnel whisk
11	bake bind board bootleg box consign crate crochet crowd deliver deposit dispatch divide dredge drill fatten feed freeze gather handle harness hatch hog load lug mail manufacture mechanize mold nail pack patch pile plow program refit repair rig roast rot salvage seal sew ship skin smuggle sort stack stink store stuff tack transfer transport unload unpack wire
12	air amputate anchor bail bang barricade bash batter beat bite blast blaze blow bluff boot bore bowl branch budge burn bust busy butt capsize case check chip choke churn clinch clog clout conk cop copy crack dabble die ditch doll dope drown fade figure fire fix flatten flood flunk fob foul grind hang hide hit keel knock lam lance land light loose map mess mow muffle muss nip overshoot pan pilot pipe plant pump punch queer rake rap reconnoitre remount rough round sally saw scram scratch settle shine shut signal slug smack smell sniff snip snitch sock souse spank spell sprain spray squash stake stamp steer stick sting stomp strike string suffocate swear tag tangle tar thaw thrash tickle time tinker tire touch tune wear whip wind winnow work wring
13	abuse accuse arrest bait balk behave benefit betray bleed boss bully butcher castrate cheat club commit deceive defraud demoralize discharge discipline disobey educate electrocute exploit fight fool humiliate imprison indulge inflict kidnap lynch mislead mistreat murder mutilate nag outnumber patronize penalize persecute pet pit plunder pollute pressure proscribe punish quarrel rape reprimand repulse revile rob scalp scold slaughter spirit starve station subject sue suffer sunburn tamper tolerate torture train trample trap treat trespass vaccinate violate waste
14	bathe boil brush bury clean clear comb cook darken dirty drain drip dust empty fill finish freshen heat lick massage mend mix polish rub scour scrub shave soak stain tidy warm wash wet whiten wipe

15	advertise barter bet borrow build buy cash charter cost donate eat hire inherit inject install invent invest migrate order own pawn pay purchase rent sell send sign spend steal subscribe team trade
16	accomplish adopt aim appoint assign assist attempt commence complete conduct consent dedicate devote edit embark encounter engage entrust fail finance incur initiate involve launch manage organize perform proceed propose prosecute publish pursue risk stage succeed superintend supervise tackle undertake venture volunteer wage
17	address adjourn appear arraign arrive attend boycott broadcast cancel canvass chair clash depart disband dissolve draft end happen issue last lodge postpone preside quit rain re-elect reconsider reconvene register release reopen request resign resume retire schedule submit subpoena suspend terminate transact transpire vacate vote waive
18	bat breakfast broil bunk burrow campaign carouse chase chew dine doze dream drink fish fry fuss hole hunt knuckle lie linger moon nap nurse perch purr rehearse relax repose rock rummage rustle scout sleep smoke snivel snooze snow sulk sunbathe sweat tote undress wake wallow wink
19	account administer afford allow arrange become compete contain cover create deal design develop establish fit form furnish include introduce lend list mark net occupy offer operate pose prepare present provide recommend remain represent require reserve reveal serve share supply support yield
20	braid brim carve celebrate color compose crown decorate display dress erect flaunt frame inhabit line ornament paint paste pave photograph roof shadow shape sketch star trim whitewash
21	acknowledge adhere admit announce argue assert assume challenge charge claim complain conclude confess confirm contradict declare demand deny exercise express insist maintain manifest oppose presume proclaim pronounce prophesy prove quote report restate state suggest testify urge
22	act base bear blame collaborate concentrate confine count depend employ err experiment focus gamble levy owe prevail rely reside resort rest specialize speculate thrive trust use
23	advance alter boost change complicate compound correspond culminate decline deepen devalue double enlarge exceed expand extend grow improve increase lower measure mount parallel peak quicken raise reduce reinforce rise slow space thicken
24	achieve acquire apply appropriate ascertain circulate collect compile deduce deduct derive dispose distribute earn effect extract gain grant locate obtain procure produce receive record recover reproduce retain score secure substitute transmit win
25	bubble cheer clap cough croak cry curse erupt explode faint holler howl hug jabber laugh mew plead pound rage ring roar scream shout shriek

	sing sneeze sound squawk storm swoon thunder vomit weep whisper whistle yell
26	add agree answer assent blurt bluster comment continue counter echo exclaim inquire murmur nod object protest react rejoin remark remonstrate repeat reply retort return smile sneer
27	accompany affect alarm amuse astonish attract bother cloud confuse depress dictate disappoint disturb excite frighten hurt impress influence interfere interrupt intrude miss offend please rouse scare shock stir surprise trouble unhinge upset worry
28	accept acquaint appreciate believe belong describe discover excuse exist explain fear feel forget guess heed ignore interpret judge learn materialize mention mind notice penetrate realize recognize regret remember respect satisfy solve stop suit suppose suspect understand unravel wonder
29	aid animate annex appeal ascribe assassinate avenge bargain brand command compel compromise condemn condescend confront conquer consolidate contrive defeat defend defy dethrone disarm disguise dismiss disprove elect elude embrace emigrate enlist evade excommunicate execute expel fend fortify govern grapple guide hail halt humor imitate impeach impose mobilize obey oblige oust outflank overcome overrule overthrow pacify persuade praise rally reason reclaim reform remake renounce revive rid rule scheme secede strive summon supersede surrender sustain tame unite unseat uphold vindicate worship wreak
30	atone balance behold belie bestow beware bid bless bloom brave breed cement comfort commend communicate consecrate consummate cooperate cultivate economize endure enforce entertain entreat exact fare forgive fulfill gloat gratify heal honor impart jest lighten minister pardon practise pray preach profit reap redeem repay requite reward rue shirk sin sow spare sup taste welcome