A corpus-pragmatic analysis of linguistic democratisation in the British Hansard
Comparing the two houses

1. Introduction

Linguistic democratisation is nowadays commonly understood as a group of discourse-pragmatic processes related to societal and sociocultural changes (Leech et al. 2009, Farrelly and Seoane 2012). In the context of English, the various manifestations of democratisation have been discussed both from a more general perspective (Leech et al. 2009) and in register-specific studies ranging from written scientific language (e.g., Seoane and Loureiro-Porto 2005) to radio chat show speech (Smith 2020). Furthermore, some democratisation processes, such as increased gender equality, have been examined from a linguistic perspective by studying the use of gendered vs. non-gendered pronouns, for example (Paterson 2014; see also Author et al. 2020a). While democratisation, colloquialisation and related phenomena can be understood and operationalised in different ways (see Mair, this volume; Author et al. 2020a), there is now a convincing body of evidence in support of the generalisation that, in the context of English, language use has become increasingly democratic, colloquial, and informal in recent history. At the same time, it is well known that linguistic variation across registers is considerable, and it is therefore to be expected that the diffusion of linguistic changes related to democratisation is also mediated by registers (e.g., Biber & Gray 2012).

One specific type of discourse where democratisation processes hold particular interest is parliamentary discourse, and this aspect has received increasing attention in a number of recent studies using the Hansard Corpus (e.g., Spirling 2016; Archer 2018; Author et al. 2020b; Author et al., forthcoming). What makes parliamentary discourse interesting from this perspective is the fact that it is produced at the nexus of several different contextual dimensions
that are known to be relevant to linguistic variation in general: speech vs. writing, planned vs. spontaneous, and formal vs. informal. From a rhetorical perspective, parliamentary speech may also take very different shapes depending on the issue under debate: at times it may be highly technical and precise, redolent of scientific discourse (Auth, forthcoming), while at other times it can be personal, emotive, and overtly persuasive, even confrontational (e.g. van Dijk 2000, Ilie 2004; 2015, Osnabrügge et al 2021, Alexander and Struan 2022). Parliamentarians also consider different types of audiences when designing their turns — those present in the parliamentary sitting, their fellow party members, members of other parties, their constituents, and other members of the public, among others — which also influences their linguistic choices.

These rather unique contextual characteristics of parliamentary discourse naturally raise the question of the extent to which they actually influence the patterns of change vis-à-vis the processes of democratisation and colloquialisation. If, as has been argued, many registers of English have indeed become more informal during the 20th century, can similar trajectories be found in the diachronic study of parliamentary discourse? Based on previous work, the tentative answer is yes, although precise patterns appear to vary depending on a number of contextual factors, such as changes in the reporting conventions (Alexander & Dallachy 2019; Author et al. 2020b). Indeed, what makes drawing such large-scale generalisations difficult is the ample contextual variation within the register of parliamentary discourse, which is all the more salient when the diachronic perspective is taken into account. This variation is difficult to incorporate into a corpus-based study design, as the quantitative analysis would have to be complemented by a detailed investigation of the external production circumstances as well as a close reading of the parliamentary texts.
The aim of this paper is to provide a more detailed examination of the impact of select contextual factors on frequency data, and the extent to which this kind of data can be linked to colloquialisation in the Hansard Corpus (Alexander & Davies 2015). Accordingly, our study investigates three linguistic features which in previous research have been connected to informal style and a high degree of speaker involvement: private verbs (e.g., think, believe), progressive verb constructions, and zero that-clauses (see e.g., Biber 1988; Biber & Finegan 1997). From a frequency perspective, our main external factor of interest is the House in the two-House system of the UK parliament (i.e., House of Commons/House of Lords), and we specifically explore how this variable influences the rates at which these features are used.

To supplement the analysis of frequency data, and to address some of the concerns identified above, we also analyse part of our data from a corpus-pragmatic perspective (Rühlemann & Aijmer 2015) with the aim of complementing our “vertical” corpus analysis with a “horizontal” reading of the data. Here, we pay particular attention to a common usage in parliamentary discourse where speakers use the private verb think with a third-person subject to signal disagreement with another person (usually another MP or Lord). In other words, just as speakers use private verbs in the first person to signal their personal stance (e.g., I think), they also use them in the third person to indicate their opinion about someone else’s position (e.g., he thinks that he has answered the debate but he has not). While taking such pragmatic functions into account in quantitative corpus analyses is potentially laborious, we argue that this is valuable, and often necessary, if we wish to obtain accurate and nuanced interpretations of frequency data — which in turn are necessary for making generalisations about the relationship between language and culture.

Our main research questions can be summarized as follows. First, do we see evidence of colloquialisation in the Hansard record as measured by the three linguistic
features/constructions discussed above? Second, are there differences between the House of Commons and the House of Lords with respect to the rate and degree of colloquialisation? And finally, if we find such differences, how can we account for them?

2. Democratisation, colloquialisation, and related processes

While the term *democratisation* is used in different ways across the humanities and social sciences (e.g., Grugel 2001), in this paper we use the term to refer to a variety of discourse-pragmatic processes in the (mostly) recent history of English. This sense of democratisation has become the standard way of using the term in linguistics, where it has recently gained momentum as an explanatory factor for diachronic variation (see e.g., Fairclough 1992; Leech et al. 2009; Farrelly & Seoane 2012; Author et al. 2020c; Smittberg 2021; Mair this volume). This work builds on the premise that as societies become more democratic, this is also reflected in language use, among other things as an increasing acceptance of informal language in contexts that have traditionally been dominated by formal and regulated usage. Precisely how this takes place, as well as what linguistic features it applies to, varies across contexts and time periods and has been subject to a lot of empirical work, often using corpus-linguistic methods. Many studies have also investigated the dynamic relationship between social change and linguistic change using related, often partially overlapping terms, such as *informalisation*, *conversationalisation*, *personalisation*, *popularisation*, and *colloquialisation*. (e.g., Leech & Smith 2009; Mair, this volume).2

---

1 Overt markers of power asymmetry may of course be substituted by covert ones (Fairclough 1992: 2003), and Mair (1997, this volume) accordingly attributes colloquialisation to “informalisation and (pseudo-)democratisation affecting advanced industrial societies” (Mair 1997: 198, emphasis added).

2 Changes in the opposite direction, i.e. towards increasingly dense expression through complex noun phrase structures and non-finite clauses, have also been identified in some genres like academic prose (Leech & Smith 2009, Biber & Gray 2011, Smittberg 2021).
Our main focus in this paper is on the last of these terms, *colloquialisation*, which is often regarded as a sub-process of discursive democratisation, although different studies use slightly different definitions and operationalisations (see Farrelly & Seoane 2012; Mair, this volume; Hiltunen et al., this volume). Following the introduction of the term, colloquialisation has been identified in numerous studies as a factor contributing to changes in the frequencies of specific linguistic features in the history of English (e.g., Hundt & Mair 1999, Leech & Smith, 2009; Leech et al. 2009; Smitterberg 2021). The basic sense of *colloquialisation* refers to changes in the norms of written language towards increasing acceptance of features that are associated with spoken usage. The original formulation is situated in the study of twentieth-century English, but more recent work has applied the notion to the study of English in the previous centuries (e.g., Biber & Gray 2012; Smitterberg 2021; Seoane & Loureiro-Porto, this volume).

As observed by Mair (this volume), the large volume of research on colloquialisation owes much to its relatively straightforward operationalisation for corpus analysis. The standard procedure has been to identify specific linguistic features which are associated with informal/spoken communicative situations — for example progressive verb forms — and trace their frequencies diachronically in genres which, based on text-external criteria, are situated closer to the formal/written end of the continuum. If an increase in the frequency of these features can be attested, this can be taken as evidence of colloquialisation (or, alternatively, of popularisation, drift, or stylistic levelling) in the relevant genres (Biber & Finegan 1989; Hundt & Mair 1999; Leech et al. 2009; Leech & Smith 2009; Rühlemann & Hilpert 2017; Säily et al. 2017; Schützler 2020, among others).

---

3 Smitterberg (2021: 91) cites Siemund (1995) and Mair & Hundt (1995) as the first studies using the term *colloquialisation*. 
An important corollary of this approach is that even if colloquialisation is defined with respect to the norms of speech and writing, studying it does not depend on the availability of spoken data from the past, as long as a link between specific linguistic features and informal spoken registers can be established, for example with the help of PDE data (Smitterberg 2021: 91). Furthermore, the term *colloquialisation* can be applied to the analysis of spoken public discourse (e.g., parliamentary speech) in particular, where an increase in informal features may be indicative of the democratisation (or pseudo-democratisation) of discourse in Fairclough’s (1992) sense (see Leech et al. 2009: 49).

With that said, it is clear that colloquialisation — and the related processes — also lend themselves naturally to a register-based investigation. While the area of register studies abounds with different definitions and understandings of *genres, registers* and *text types* (e.g., Ferguson 1994; Biber & Conrad 2009), there is a general consensus that the register perspective provides a useful way of conceptualising and compartmentalising the vast universe of texts and discourses for diachronic linguistic analysis. Thus, by taking as the point of departure the same process of societally-motivated linguistic change, and contrasting its progress across registers in real time, it is possible to capture one aspect of register evolution. For example, such comparisons have enabled Hundt & Mair (1999) to characterise newspaper writing as an “agile” register of written English: linguistically, it is easily influenced by a growing trend towards informality and communicative immediacy in a way that a more “uptight” register, such as academic prose, is not. In the same vein, Biber & Finegan (1989) have identified a systematic pattern of change affecting some written genres, which consists of successive and gradual linguistic developments in the same direction, namely towards a more oral style. Regarding these types of analyses, Biber & Gray (2013) have argued that it is important to adopt operationalisations of register categories that are maximally specific and accurate, as it is at the level of such registers that linguistic changes are mediated, rather than broader registers.
where the changes are conflated. From this it follows that if the objective is to describe changes in, for example, newspaper language, then studying the language of newspaper editorials, classified ads and sport news separately is likely to prove more fruitful than focusing on the register of newspaper writing at large (cf. Author 2021).

The register of parliamentary discourse provides an interesting testing ground for a register-sensitive study of colloquialisation. First, it represents public discourse related to governance, administration and politics, where issues of democratisation are naturally of particular interest. Second, corpus data is available for this register in the form of the Hansard Corpus, which covers a period of over 200 years and thus enables the study of real-time linguistic changes and their potential connection to social change. Third, the register itself is situated at the intersection of spoken and written discourse, and as such, it is subject to the influence of the norms of both spoken and written language, at least in principle. Technically, it represents spoken discourse (insofar as the term refers to speech delivered in parliament), which is often planned and put down in writing before the actual delivery. After the delivery, parliamentary speech is converted into written transcripts, which make up the Hansard record and on which corpus-based studies of parliamentary discourse are based.4

Previous studies of the Hansard have explored many themes relevant to democratisation and colloquialisation. For example, based on an analysis of sentence length and word length, Spirling (2016) argues that parliamentarians’ utterances became less complex and easier to understand after the electorate had doubled from one to two million men in the Second Reform Act in 1867. Archer (2018) studied the use of politeness features (e.g., with ... respect) diachronically, showing that their incidence in the Hansard is related to a complex interplay of

4 A well-known caveat in the historical study of Hansard is that the transcriptions do not provide a fully accurate representation of the parliamentary debates. As discussed in Mollin (2012), the transcribed texts include omissions and amendments, and the editing principles of the Hansard texts have also changed over time.
factors involving not only facework behaviour and specific discourse topics, but also reporting
conventions and ritualistic modes of address to negotiate differences. Hou and Smith (2018)
investigated the decline of passivisation (characteristically a feature of formal style), while
Author et al. (2020b) found evidence of a colloquialisation trend specifically (though not
exclusively) related to the introduction of the so-called *Official Report* in the first two decades
of the 20th century.  

However, previous work has rarely paid attention to potential differences between the two
Houses of the Parliament, the House of Commons and the House of Lords (see, however,
Archer 2018). These Houses differ in terms of their make-up, membership and mandate. For
example, while the members of the House of Commons are elected for a five-year term in a
general election, most members of the House of Lords are either appointed for life (“life peers”)
or have inherited their membership (“hereditary peers”). Furthermore, the members of the
Commons engage in a variety of tasks, including the preparation and passage of new
legislation, funding decisions, and the examination of government policies, for example, while
the work in the House of Lords focuses on law-making in its various stages. From a political
perspective, the great majority of the MPs in the Commons are affiliated with a political party,
whereas fewer than half of all the Lords are politically affiliated (politically neutral members
include the “Lords Spiritual” (bishops), crossbenchers, and non-affiliated Lords). From the
perspective of age and gender, the average age in the House of Commons is considerably lower
than in the House of Lords. After the 2019 general election, the average age of the newly
elected MPs in the Commons was 51, while according to a 2017 factsheet, the average age in
the Lords was 69. Finally, although women are in the minority in both Houses, they are better
represented in the House of Commons, where 34% of all members are women; the proportion

---

5 The early Hansard record represents third-person narratives compiled from newspaper reports, and the so-called
“verbatim record” was only introduced in the first decades of the 20th century. This change in the production
circumstances is also manifested as clear linguistic changes in the parliamentary record (Author et al. 2020b).
of women in the House of Lords is only 28% (according to 2019 data; Uberoi et al. 2022: 6). Given these external differences between the House of Commons and the House of Lords, it is reasonable to treat the speeches delivered in them as different sub-registers of parliamentary discourse. Our aim in this study is to investigate empirically to what extent these external differences are manifested in language use vis-à-vis features associated with colloquialisation.

3. Features in focus

3.1 Private verbs

Private verbs have been studied extensively in previous research, and they have been investigated from various standpoints. From a syntactic perspective, many private verb patterns, such as *I think*, *I guess* and *I suppose*, can be used in different positions in the clause, and while they typically occur clause-initially (e.g., *I think you’re correct*), they can also be used parenthetically in a clause-medial (*what you said is, I think, correct*) or clause-final position (*what you said is correct, I think*). These parenthetical uses are in many ways similar to adverbial usage (e.g., *what you said is correct, perhaps*), which is why some researchers have proposed that epistemic/evidential patterns like *I think* should be regarded as grammaticalised “epistemic phrases” or “pragmatic markers” (see e.g., Thompson & Mulac 1991; Palander-Collin 1999, Denis 2022). The grammaticalisation argument gains further currency from the observation that the verbs in these patterns show signs of desemanticisation: instead of expressing psychological states or processes denoted by the verb, they are more often used in an interpersonal function as expressions of epistemic/evidential stance (see e.g., Kärkkäinen 2003, Van Bogaert 2010).

---

6 Also referred to as “cognitive verbs”, “psychological verbs”, or “verbs denoting mental processes” in different studies (Fetzer 2014: 70).
The interpersonal function of private verbs has also received attention in register studies. Biber (1988) demonstrated that the use of private verbs with first- and second-person pronouns correlates with a high degree of speaker involvement, which is typical of oral communication, and of informal conversation in particular. Importantly from the perspective of the present paper, these involvement features have also been identified as relevant components in political discourse. For example, Fetzer observed a diachronic increase in the frequency of three epistemic parentheticals (I mean, I think and I believe) in recent political speeches and pre-election interviews with Labour, Tory and Liberal Democrat politicians, which she interprets as supporting “the critical-discourse-analytic claim of an ongoing process of conversationalisation in British institutional discourse” (2014: 78). While our research is framed in terms of colloquialisation instead of conversationalisation, one of the aims of this study is to determine the extent to which this trend can be detected in another register of political discourse, i.e., parliamentary debates.

In this study, we first examine private verbs from the perspective of colloquialisation. We start by investigating their frequency in the House of Lords and the House of Commons in order to see whether there is variation in the discourse styles between the two Houses. In our quantitative analysis, we focus on first-person usage for three reasons. First, as discussed above, first- and second-person forms have been argued to be particularly indicative of high speaker involvement, informal style, and conversationalisation (Biber 1988, Fetzer 2014). Second, members of parliament are generally prohibited from addressing one another in the second person during parliamentary proceedings (Factsheet G7: 2), which means that second-person usage is both highly regulated and scarce in general; for this reason, we have excluded second-person forms from the analysis. Third, while going through our data qualitatively, we identified a third-person usage of the most frequent private verb think, which shows frequency changes both in the House of Lords and in the House of Commons, but which does not seem
to be directly linked to colloquialisation. In this pattern, the referent of the third-person subject of the matrix clause is coreferential with the subject of the complement clause (e.g., he_1 thinks that he_1…), and the pattern is often used for particular pragmatic effect: in contrast to the first-person pattern, through which speakers express their opinion or degree of certainty, the third-person coreferential pattern is often used to imply disagreement with the position previously expressed by the referent of the third-person pronoun, as in (1) and (2). The source of disagreement is typically retrievable from the surrounding discourse.

(1) First of all, the hon. Member is mistaken if he thinks that he has a right to ask a supplementary question. (Commons, 1952, William Morrison)

(2) It is a fact that the Minister is flouting the advice he is receiving and has received from every quarter. He thinks that he knows best. I understand that it has been said of him that once he has written out a political prescription he will claim that he is right even though the patient dies. (Commons, 1961, Kenneth Robinson)

In (1), the speaker explicitly states that the MP whose position he refers to is mistaken in his assumption that he has a right to ask a question (the proposition expressed in the complement clause), while in (2), the source of disagreement is more broadly expressed in the preceding and following discourse. As presented, our intention is not to analyse this pattern from the perspective of colloquialisation, but rather to emphasise the point that corpus-based studies must always consider the possibility that the data may include trends and developments that are difficult to connect directly to the explanatory model employed in the study. In our case, even if private verbs can in general be reasonably linked to colloquial style, the frequency of this particular pattern, and any changes therein, must probably be explained in other ways. Therefore, as far as private verbs are concerned, it is prudent to restrict the investigation of colloquialisation to first-person forms.
3.2 Progressives

The second feature in focus, the progressive, is likewise a convenient index for the degree of colloquialisation and has also been treated as such in previous studies (see Smitterberg 2021: 90). In Present-day English, the progressive is clearly more associated with informal language and spoken discourse than with writing (Biber et al. 1999: 461–463; Leech et al. 2009), but its frequency in writing has been on the increase both from a centuries-long diachronic perspective (e.g., Hundt 2004; Smitterberg 2005; Kranich 2010) and in recent British English and other varieties of English (e.g., Smith and Raeyson 2007; Kirk 2015: 88; Rautionaho & Fuchs 2020).

According to the data from ARCHER presented in Kranich (2010: 107), the increase started in the seventeenth century and had not plateaued by the early twentieth century despite having slowed down. Leech et al. (2009) show that the frequency of progressives was still increasing in the twentieth century both in printed prose genres and in speech; the former finding supports the colloquialisation hypothesis (written registers adopting speech-like features) and the latter can tentatively be attributed to the grammaticalisation of new functions of the progressive (Leech et al. 2009: 126, 142). In a recent overview, Mair (2021) concludes that the spread of the progressive has still not reached its saturation point in either speech or writing and that the main determinants of variation are genre/register and style, rather than time and geographical region. In particular, he notes that the overall increase in the frequency of the progressive is not primarily driven by its use with stative verbs — e.g., *I’m liking this* — but is rather a compound effect of the increased frequency of different functional extensions (e.g., passive progressives, interpretative uses, and conventionalised prefabricated units like *I’m just saying*) (Mair 2021: 774–776).
However, the overall patterns of development of the progressive are not only complex with regard to register but also contingent on a variety of lexicogrammatical patterns and collostructional preferences (Rautionaho & Fuchs 2020), and this needs to be kept in mind when using it as a measure of colloquialisation (see also Smitterberg 2007: 282). Another difficulty related to a corpus-based analysis of progressives is the fact that the accurate identification and classification requires manual analysis, which means that some compromises in precision and recall are inevitable when the study is based on a large corpus that does not include syntactic parsing, such as the Hansard Corpus. Despite these caveats, the frequency of the progressive remains a useful indicator of colloquialisation together with the other two features used in this study (private verbs, that-deletion).

3.3 That-deletion

The third linguistic feature that we will study from the perspective of colloquialisation is connected to our first topic of interest, i.e., private verbs. We examine the use of the complementizer that with three types of verbs — private verbs, public verbs, and suasive verbs — paying particular attention to diachronic trends in the retention of that and its omission (“that-deletion”). That-deletion has been extensively studied in the past, and it has been connected with informal style and speaker involvement in numerous studies since the 1960s (see Bryant 1962: 209; McDavid 1964: 113; Storms 1966: 262–264; Elsness 1984; Aijmer 1997: 10; Biber 1988; Kaltenböck 2006). The strong connection between that-deletion and informal style makes it an excellent feature to study from the perspective of colloquialisation, as the predictions concerning its occurrence are clear-cut: the more often we find that omitted in complement clauses governed by these three types of verbs, the more colloquial the style of the text. However, the analysis of the use of that is complicated by the fact that in order to
determine the true rate of its occurrence, it would be necessary to identify all contexts where *that* could have been used in the data but was not. Given that this would require the researcher to go through the entire corpus manually, this is typically not feasible, at least when the data comprise big-data corpora such as the Hansard Corpus.

In addition to being associated with informal style, *that*-deletion has sometimes been argued to reflect meaningful differences when compared to a clause where *that* is overtly expressed (Bolinger 1972; Underhill 1988; Dor 2005). In Underhill’s data, for example, *that* was commonly deleted when the author of the text endorsed the assertion in the complement clause, whereas *that* was more often retained when the author did not endorse the assertion. While we will only analyse the bulk of our data more broadly from the perspective of informal style and colloquialisation, we will keep Underhill’s analysis in mind when examining our data on the third-person coreferential construction of the type discussed in section 3.1 (e.g., *he thinks that he has a right to ask a supplementary question*).

4. Material and methods

4.1 Hansard corpus

All our case studies in this paper make use of the 1.6-billion-word Hansard Corpus. Based on data from the historic Hansard, the corpus contains “nearly every speech given in the parliament from 1803–2005” (Alexander and Davies 2016) and various levels of annotation (including tokenisation, POS-tagging, lemmatisation, and semantic tagging). The quantitative analysis represents the entire time period covered by the corpus and comprises all material from both houses of Parliament.
4.2 Data retrieval and analysis

As presented in Section 3, we focus on three linguistic items/phenomena that in previous work have been associated with colloquialisation. As we are working with a large corpus, we have adopted a search strategy whereby we make use of POS-tags to retrieve general frequency patterns, which we then use to generate time-series plots for visual exploration. Our aim is to strike a good balance between feasibility and acceptable levels of precision/recall. In other words, we are prepared to accept less-than-optimal rates of accuracy in the data set, assuming that this limitation will partially be compensated for by the amount of data.

To obtain frequencies of **private verbs**, we used the list of 67 private verbs in Quirk et al. (1985: 1181) complemented by the verbs listed in Biber (1988). We determined the frequencies of private verbs with first-person singular subjects (e.g., *I think*), as these are particularly useful as indices of colloquialisation. The frequencies of the first-person forms are based on the bigrams [*I + private verb*], which would seem to offer a reasonably accurate approximation of the general diachronic trend despite the fact that, e.g., negative clauses (e.g., *I don’t believe*) and instances with an intervening adverb (e.g., *I really think*) are left out from the dataset.\(^7\)

To retrieve **progressive forms**, we searched for combinations of forms of the verb *BE* and the *ing*-forms of verbs, using the following sequences of POS-tags, where VVG stands for the *ing*-form:\(^8\)

- [VB0] [VVG]
- [VBDR] [VVG]
- [VBDZ] [VVG]

---

\(^7\) In this specific example, the token frequency of the bigram *I think* in the entire *Hansard Corpus* is 1,357,503, whereas the token frequency of *I *_<R*> think* (where <R*> stands for any adverb) is over 50 times smaller at 25,673.

\(^8\) See http://ucrel.lancs.ac.uk/claws7tags.html for the full tagset.
As with private verbs, these queries do not capture all progressives (e.g., negative sentences or sentences with intervening adverbs are excluded), but similarly to our private verb data, we regard the dataset compiled with these queries as representative of general trends concerning the use of the progressive. Furthermore, due to the amount of data, we did not distinguish between true progressives and be + participial adjectives (e.g., *the show was entertaining*; Huddleston and Pullum 2002: 80), and be + gerund (e.g., *To do so is taking a prudent step*; Smitterberg 2005: 26), as this would require extensive manual post-processing of the data. Even so, spot checks indicate that the amount of noise in our data is negligible, and we consequently consider the results obtained by our corpus queries to be indicative of the overall frequency of the progressive construction.

With *that*-deletion, we adopted a strategy similar to the one in Biber (1988) and limited our focus on specific verb types that license *that*-clauses: private verbs (e.g., *think, believe, and guess*), public verbs (e.g., *exclaim, pronounce, and declare*), and suasive verbs (e.g., *agree, demand, and suggest*). To ensure that our data accurately represents the variation between the use and non-use of the complementiser, we further narrowed down the scope of our analysis to sequences where the verb is followed by a pronoun, an optional complementiser and a verb or an auxiliary (e.g., *think (that) he knows*). These queries, too, only capture part of the relevant sentences, but they are treated as representative of contexts where complementiser variation (*that* vs. zero) may occur. Using this proxy, we determined the proportions of sentences with
and without an overt complementiser. We also decided to study each private verb individually because there is no reason to assume that the diachronic trends would be similar for every verb. Furthermore, if we studied all verbs as a single group, we would not be able to examine the changes in the use of low-frequency verbs, as the data for some high-frequency verbs (most notably think) would dominate the results.

After determining the frequencies of these features in the corpus separately for each decade and each house, we created visualisations of the time series for exploratory data analysis in order to identify any notable changes in the frequencies across the time period in focus. The quantitative analysis is complemented by a qualitative analysis where we consider some possible explanations for the observed trends.

5. Results

5.1. Private verbs

For all three selected features, the colloquialisation hypothesis predicts that the frequencies should show an upward trend that proceeds in tandem with increased societal democratisation and the gradual informalisation of manners and linguistic behaviour. For many private verbs, the predicted pattern can indeed be attested. This can be observed in Figure 1, which shows the normalised frequency of all first-person singular pronoun + private verb constructions, with the frequencies of the five most frequent verbs highlighted separately (think, hope, believe, understand, know).

The data show an upward trend both for the aggregate frequency of these private verb constructions and for four of the five high-frequency verbs studied: believe, hope, know, and understand. However, the data on think, the verb with the highest overall frequency in our private verb dataset, yields contrary results: we first see a substantial frequency increase in both
Houses, which is then followed by an equally dramatic decrease in the second half of the twentieth century. This can probably be explained by differences in editorial conventions, which we will discuss more in Section 6.

**Figure 1.** Time series of select private verbs (normalised to 1,000,000 words).

The data represented in Figure 1 are also noteworthy from the perspective of subregister: we can see that while both Houses follow the same overall trends, there are also differences in the rate of the development. In the House of Commons (left panel), the frequency of private verb constructions increases steadily from the early 1800s to the first decade of the 20th century (1800s: 610 pmw; 1900s: 1,600 pmw), after which there is a particularly marked frequency increase in the 1910s (3,400 pmw), followed by a plateau. The trend looks largely similar for the House of Lords (right panel), but the rapid increase in the frequency of private verb patterns takes place two decades earlier than in the House of Commons: the frequency of private verb constructions in the 1890s is c. 3,200 pmw, which is twice as high as in the previous decade.
This observable difference between the two Houses can in large part be attributed to the changes in the reporting conventions, and these changes were in turn clearly motivated by concerns about democratic and transparent public administration. While the early Hansard reports were compiled from newspaper accounts, which were mostly third-person summaries of parliamentary speeches, these were replaced in 1909 by the *Official Report* in the House of Commons, which consisted of “substantially verbatim” records of the speeches (Author et al. 2020b: 3–4). However, a similar system had already been introduced in the Lords in 1889, owing to which the material that is available for the House of Lords from 1890–1910 contains a much fuller representation of the actual speeches than the corresponding sections in the Commons (Jordan 1931: 442).

While differences in the reporting conventions are relevant to the frequencies of private verbs in the two subcorpora, it does not appear that there are any functional differences in private verb usage in the two Houses: the first-person patterns are used to signal the speaker’s personal opinions and beliefs in both the House of Commons and the House of Lords, as in (3) and (4).

(3) I *find* this argument unconvincing and rather pathetic, and I hope that my hon. Friend will think again. (Commons, 1986, David Atkinson)

(4) I *believe* that we should not accept this clause. (Lords, 1980, Lord Butler of Saffron Walden)

5.2. Progressive verb phrases

For *progressives*, the aggregate data exhibits a clear diachronic increase from 1850 onwards, which would again seem to match the prediction of the colloquialisation hypothesis. Between 1850 and 1900, the frequency of progressive verb phrases quadruples, and the growth pattern
accelerates even further in 1900–1920, after which the frequency hits a plateau at around 2,700 pmw.

Figure 2. Frequency of progressives in the Hansard Corpus (aggregate).

However, if we again look at the data from the two houses separately, we notice that the frequency of progressives has changed at different rates. This time, the House of Commons exhibits higher frequencies of progressives in the earlier decades, as well as generally somewhat higher frequencies in the 20th century, after the introduction of the *Official Report*.9

---

9 The apparent peak in the final bar (corresponding to years 2000–2004) is likely to be a corpus artefact, as this section of the corpus is much smaller than the sections representing the earlier decades.
The data also show that the frequency of the auxiliaries that are most strongly connected with high speaker involvement, *am* and *are*,\(^\text{10}^\text{10}\) show a particularly pronounced increase in the early twentieth century in the House of Commons, although *is* increases in frequency as well. Jointly, *am* and *are* comprise nearly half of the data in the most recent decades. Here, it should be noted that the increased frequency of *are* is primarily a consequence of the frequent use of first-person plural forms; Figure 4 compares the frequency of progressives with pronominal subjects in the first- and third-person plural (i.e., *we are* V-Prog vs. *they are* V-Prog) in the entire corpus. The data show that progressives in the first-person plural are not only more frequent than third-person forms but also that their frequency has increased quite substantially in the

\(^{10}\) As discussed above, first- and second-person pronouns have been shown to be particularly frequent in interactive and involved speech production (e.g., Biber & Finegan 1997). By extension, first- and second-person auxiliaries can also be argued to indicate high speaker involvement.
twentieth century (from 256 pmw in the 1910s to 453 pmw in the 1960s), while the frequency of third-person forms has stayed roughly the same. Examples (5) and (6) illustrate typical usage.

(5) As the result of the way in which collieries have been mismanaged, we are seeing the starkest tragedy overwhelming communities that have no hope of any other kind of work than coal-mining. (Commons, 1938, Stephen Davies)

(6) When we are considering age exemption, and whether or not the limit for age exemption should be raised, let us remember that there was no provision for age exemption at all in our tax system before my right hon. Friend the Minister of Aviation introduced it in 1957. (Commons, 1961, Sir Edward Boyle)
Figure 4. Frequency of progressive verb phrases in the first- and third-person plural (aggregate).

5.3. Complement clauses and *that*-deletion

The relatively steady increase in the frequency of private verb patterns and progressive verb forms provides initial support to our hypothesis according to which parliamentary discourse has become colloquialised over time. However, our data on private verbs that are followed by a complement clause yield somewhat conflicting results. As presented in Section 3.3, the omission of the complementizer *that* has widely been regarded as a feature of informal or colloquial speech, and the colloquialisation hypothesis would accordingly predict that *that*-deletion should become increasingly frequent in the more recent periods under study. However, as shown in Figure 5, we see a more mixed trend. While there are many private verbs where the proportion of *that*-deletion has remained largely unchanged (e.g., *add, agree, state, suggest*), and there is one verb (*allow*) that conforms well with the colloquialisation hypothesis, many verbs show a marked increase in the retention of *that* (e.g., *believe, hope, consider, say, think*).
Figure 5. The proportion of *that*-deletion in select private verb constructions.

Interestingly, Figure 5 not only shows that the data on *that*-deletion does not support the colloquialisation hypothesis but also that there are no substantial differences between the two Houses: the observed trendlines show considerable overlap, and while for some verbs (*believe, hope, think, understand*), a declining trend in *that*-deletion starts a decade or two earlier in the House of Commons than in the House of Lords, the datasets largely converge in the most recent periods studied. We will come back to potential explanations for this development in Section 6 but note here that the changes in the proportion of *that*-deletion are in some cases quite remarkable. For instance, up until the 1950s, nearly 80% of the complement clauses after the high-frequency verb *think* occurred without *that* in both Houses. In the 2000s, by contrast, the proportion of *that*-deletion has plummeted to 23% in the House of Commons and to 36% in
the House of Lords. Such a dramatic change raises suspicion of editorial interference in the reproduction of parliamentary discourse in the late nineteenth century.

5.4. The pragmatics of coreferential complement clauses

We conclude our study of the linguistic democratisation of the House of Commons and the House of Lords with a qualitative case study of a particular function of the most frequently occurring private verb in our data \((\text{think})\). In our close reading of the data, we identified a rather consistent trend where the members of both Houses used \text{think} to imply disagreement. This implication is triggered in contexts where the third-person subjects of the main clause and the complement clause are coreferential, as in (5) and (6).

(5) I do not know what the Minister means by a public corporation. He has not said that its property is vested in the Minister, like the shares in BL, so who does he claim owns it? \textbf{He thinks that he has answered his own question, but he has not.} (House of Commons, 1986, Mr. Robyn Maxwell Hyslop)

(6) \textbf{If he thinks that he will get away with describing it as “a largely technical measure”, he has another think coming.} This is a thoroughly unsatisfactory order: It is self-contradictory. (House of Lords, 1994, Lord McIntosh of Haringey)

In order to mitigate the potentially face-threatening effect of the utterance, the pattern is frequently expressed in an \textit{if}-clause, as in example (6). In the House of Commons, the pattern was expressed in an \textit{if}-clause in 46\% of all cases (N=117) when the MP disagreed with another MP. The corresponding proportion for the House of Lords dataset was 48\% (N=14). Interestingly, when an MP disagreed with someone outside the House of Commons (e.g., their constituent), \textit{if}-clauses were not used at all (0 out of 12).
Figure 6 shows that the usage is common in both Houses, although the exact proportions differ somewhat. In the House of Commons, the pattern is used to indicate disagreement 57% of the time, whereas in the House of Lords, the corresponding proportion is 47%. However, the dataset for the House of Lords is much smaller than that for the House of Commons (79 tokens vs. 477 tokens, respectively), which may in part explain the discrepancy. Indeed, while this difference may on closer scrutiny point to a genuine difference in the discourse styles of the two Houses, insofar as expressions of disagreement are concerned, from a quantitative perspective the result is not statistically significant ($p = 0.08$; Fisher’s exact test, two-tailed).

Figure 6. The coreferential *he thinks that he...* pattern: disagreement vs. neutral use in the House of Lords (upper panel) and the House of Commons (lower panel).
A closer examination of the data shows that this usage is indeed situation-specific in the sense that the MPs and the Lords use it primarily as a device for debate: the vast majority of the uses in both Houses can be connected to debates where the speaker expresses a difference of opinion with another MP/Lord instead of someone who does not participate directly in the parliamentary proceedings. The results where the type of referent is taken into account are visualised in Figure 7. In cases where the referent of he is a colleague (i.e., another member of the House), 69% of the tokens are used in a context where the speaker disagrees with the referent of he in the House of Commons, while the corresponding proportion in the House of Lords is 58% (again, the Lords dataset is much smaller than the Commons dataset). By contrast, when the referent of he is not an MP or a Lord, the pattern is typically used neutrally to summarise the referent’s opinion: the pattern is used to imply disagreement only 15% of the time in the House of Commons and 21% of the time in the House of Lords. Examples implying disagreement were already illustrated in examples (3) and (4) above, while examples (7) and (8) illustrate neutral usage.

(7) … at first sight it seemed to me to be rather difficult to say that a workman in the north shall not come down to the south if he thinks that he has a chance of obtaining employment there which he cannot obtain in his own district. (Commons, 1933, Neville Chamberlain)

(8) This amendment would remove the right of the head of the local authority’s paid service to claim the resources which he thinks that he needs for his functions under Clause 4 and leave him dependent on what the local authority thinks is sufficient. (Lords, 1989, Lord Hesketh)
Figure 7. Disagreement vs. neutral usage according to the type of referent of he (MPs/Lords vs. others): the House of Commons and the House of Lords.

We also analysed our data from a diachronic perspective in order to see whether the frequency of this particular pattern has changed over time and if there are any differences between the two Houses. As discussed in section 3.1, members of parliament are not generally allowed to address one another in the second person, which makes third-person patterns like these the most direct form of address permitted in parliament. From the perspective of colloquialisation, however, this pattern does not seem to represent a style of speaking that would be particularly informal or colloquial. Considering this, it is interesting to note that even in this case the frequency of the pattern changes very differently in the two Houses. Figure 8 provides a visualisation of all the cases where the speaker uses the pattern to indicate disagreement. We
can see that while the frequency of this pattern increases steadily in the House of Commons, the Lords data show a contrary trend.

![Figure 8](image)

**Figure 8.** Frequency of *he thinks that he*... in the House of Lords and the House of Commons. Hansard Corpus. Normalised to 1,000,000 words.

Finally, we also examined whether *that*-deletion might have an impact on the pragmatic pattern noted above, bearing in mind Underhill’s (1988) suggestion according to which the loss of *that* may correlate with speaker endorsement. For this purpose, we collected 100 tokens of a pattern where *that* was omitted (*he thinks he*...) and analysed the data according to the referent of *he* and type of usage (neutral vs. disagreement) as before.\(^\text{11}\) As our dataset on *that*-omission is much smaller than our main dataset and considering that the main dataset did not yield

---

\(^{11}\) For this purpose, we randomly chose a single decade (the 1980s) for the basis of our analysis. The dataset included a total of 101 hits of the pattern *he thinks he*. One of the tokens turned out to be a false positive, and the remaining 100 tokens were consequently analysed from the perspective of agreement vs. disagreement.
conflicting results for the two Houses, we decided to pool the data from the House of Commons and the House of Lords together and did not analyse the data from the perspective of subregister or time.

The results of our analysis indicate that the omission of *that* plays no role in the observed phenomenon. In all, *he thinks he* is used to imply disagreement, doubt, or difference of opinion 59% of the time when the referent of *he* is another MP/Lord (N=39), while the corresponding proportion for other referents is just 18% (N=6). In other words, while *that*-omission is surely a relevant phenomenon to study from the perspective of colloquialisation and editorial intervention, it does not seem to affect the pragmatic tendency noted above, although a larger dataset would of course allow us to establish this with more certainty.

6. Discussion

In this paper, our main goal has been to investigate the Hansard Corpus in order to see whether the gradual democratisation of British society has resulted in the adoption of a more colloquial discourse style in the two Houses of the Parliament of the United Kingdom. We tested this “colloquialisation hypothesis” by examining the text frequency of three linguistic features/constructions which in earlier research have been connected with colloquial or informal style: private verbs, progressive constructions, and *that*-deletion. Two main results emerge from our analysis: first, there is clear evidence of a colloquialisation trend in the Hansard, and second, colloquialisation proceeds at different rates in the two Houses. We argued earlier that it would be sensible to treat the texts produced in the two Houses as representing separate sub-registers based on external differences between the Commons and the Lords; the results of this study indicate that there are also linguistic differences that lend support to this distinction.
However, there are also some caveats that we wish to discuss with regard to our results and our methods. While two of the three linguistic features examined in this study (private verbs, progressive verbs) yielded relatively uncontroversial evidence in support of the colloquialisation hypothesis, the data on *that*-deletion showed more variation. Indeed, for many verbs, such as *hope* and *think*, the complementizer *that* is much more often retained in the most recent decades of the Hansard Corpus than in the earlier periods. Furthermore, although the frequencies in the two Houses generally follow the same trends, we find differences in their timing. For instance, in case of *think*, the proportion of *that*-deletion plummets from c. 84% in the 1940s to c. 33% in the 1960s in the House of Commons, while in the House of Lords, a similar trend can only be seen from the 1970s onward, when the proportion of *that*-deletion starts to go down from its peak at c. 79% to c. 36% in the 2000s.

Considering that the decrease in the proportion of *that*-deletion is rather dramatic in such a short period of time, and that the timing of the trend differs between the Houses by some decades, we suggest that this finding might be best explained by editorial norms and changes therein. Given the high frequency of private verbs such as *think* and *hope* in the corpus, it is possible that the editors of the Official Report were particularly sensitive to the colloquial flair of the pattern when the complementizer *that* was deleted, and therefore decided to use *that* in the published texts regardless of whether the speakers themselves used it or not. The difference in the timing of the trends between the Houses could also be explained by editorial practices: each House has their own Official Report with their own Managing Editors. If the editors, and the staff responsible for the two publications, work relatively independently from each other, the editorial norms of the two Official Reports may differ to some extent, and some general stylistic rules may have been introduced at different times by the editors. If this explanation is correct, it further underscores the need to consider the texts from the two Houses as distinct sub-registers in all future studies that make use of the Hansard Corpus. In this respect, it is
unfortunate that as a default, the user interface through which the Hansard corpus is typically accessed (https://www.english-corpora.org/hansard/) presents the frequency data as aggregated for both Houses.

In addition to our quantitative exploration, we performed a qualitative analysis of the pragmatics of the third-person coreferential use of *think*. The motivation for including the analysis of the pattern in this study was two-fold. First, we wished to emphasise the methodological point that a corpus-pragmatic analysis of discursive change necessitates a “horizontal” analysis of the data in addition to a “vertical analysis”. In our case, carrying out a lexeme-based analysis of *think* and other private verbs would have run the risk of overlooking important functional differences between first- and third-person constructions and confounded our interpretation of the results concerning the colloquialisation of parliamentary speech. Second, this particular use of *think*, which to our knowledge has not been discussed in previous literature, is directly connected with the pragmatics of the debates that take place daily on the floor of parliament, where the parliamentarians are forbidden from using more direct forms of address (i.e., second-person forms), and we felt that a closer analysis of the phenomenon was merited. Indeed, it may be that the argumentative style of speech in parliament made it possible for us to observe this usage in the first place. Be that as it may, it is interesting to note that, in addition to expressing their personal stance with private verbs in the first person, speakers also do so in the third person.

This constructive interplay of quantitative and qualitative analysis — a foundational principle of corpus pragmatics — also emerges in the context of specific methodological choices. Our analyses are based on subsets of easily retrievable constructions that are assumed to be representative of the register features in focus, following a common practice with large corpora. While this is not an issue for comparability as the same retrieval strategies were used for both
Houses, data with a higher degree of granularity would naturally be beneficial for the interpretation of the results. Another way to refine the present results would be to look into the reasons for individual peaks in the data, as this may give a clearer insight into the factors contributing to the use of the features under study.

In light of the above, the main takeaway of this study is to underline the fact that despite exhibiting clear tendencies towards colloquialisation, the Hansard Corpus is not a monolithic linguistic source and should not be treated as such. As far as register is concerned, we suggest that it would be prudent to treat the texts from both Houses as distinct sub-registers, given the differences both in the status and demographics of the House of Commons and the House of Lords, and in the editorial conventions in producing a record of the debates. Indeed, we need to learn much more about the external production circumstances of the Official Reports if we wish to use the Hansard Corpus as evidence of language change or socio-pragmatic processes like colloquialisation. The editors’ influence on the Hansard texts is likely to be so substantial that the evidence of colloquialisation that we see in our data is probably in no small part due to the relaxation of editorial norms. This is plausible even if there is evidence of a contrary trend for some of the features associated with colloquialisation (e.g., the retention of that with think); after all, as a register, parliamentary texts are affected by the norms of both written and spoken language, and the editors are consequently likely to pay attention to the relative formality of the texts while also allowing the representation of some speech-like features in them. The relationship between the actual colloquialisation of parliamentary speech, and the extent to which the editing of the Official Reports has followed the more general colloquialisation trend witnessed in many other written registers of English, is something that needs to be better established in future research.
References


Author (2021).

Author (Forthcoming).

Author et al. (2017).

Author et al. (2020a).

Author et al. (2020b).
Author et al. (2020c).

Author et al. (Forthcoming).


by moving from specific usage patterns to more abstract categories (e.g., lemmas)