Democratisation – how can historical corpus pragmatics contribute to understanding changes in the recent history of English?

Turo Hiltunen\textsuperscript{a}, Turo Vartiainen\textsuperscript{a} and Jenni Räikkönen\textsuperscript{b}

University of Helsinki\textsuperscript{a} and Tampere University\textsuperscript{b}

1. Introduction

Democratisation has emerged as an important topic in recent sociolinguistic and discourse-pragmatic research, particularly in English studies. The growing body of research in this area has provided evidence for a trend towards increasing linguistic and discursive democratisation, whereby linguistic features that are associated with marking power asymmetries and strong obligations are increasingly replaced by more egalitarian, less face-threatening alternatives (see, e.g., Fairclough 1992; Myhill 1995; Leech et al. 2009). Democratisation (and related processes such as colloquialisation, see below) has accordingly been cited as a factor accounting for linguistic changes across a wide variety of public discourses, including news writing and other media discourses, political discourse, as well as various specialist discourses, such as academic research writing (e.g., Hundt and Mair 1999; Mair and Leech 2006; Leech et al. 2009; Rühlemann and Hilpert 2017).

The current special issue aims to increase the current understanding of democratisation in the history of English by offering a historical corpus-pragmatic perspective on the theme. The intersection of the fields of historical linguistics, pragmatics, and corpus linguistics is in many ways a natural place to explore linguistic and discursive democratisation. The adjective “historical” in this context is of course uncontroversial because democratisation, by definition,
is a process that can only be approached from a diachronic perspective. It is equally clear that the process of democratisation can fruitfully be studied from a pragmatic vantage point, especially if we espouse the broad Continental tradition of pragmatics, which represents “a general functional perspective on every possible aspect of linguistic behaviour” (Verschueren 2013).

Recently, pragmatic enquiry has increasingly turned to corpora to obtain empirical data, leading to the emergence and consolidation of the field of corpus pragmatics, “the science that describes language use in real contexts through corpora” (Romero-Trillo 2017: 1), which has become a flourishing area of study (Rühlemann and Aijmer 2015: 23; Landert et al. 2023: 7). While the application of corpus pragmatic approaches to historical data is certainly not uncharted territory – this is evident, for example, in numerous contributions to the Journal of Historical Pragmatics and several edited collections (e.g., Taavitsainen et al. 2014; Fanego and Rodriguez-Puente 2019; Hiltunen and Taavitsainen 2022) – we argue that their potential has not been exhausted in the study of how democratisation phenomena are manifested at different junctures in the history of English. In particular, we envision that the following areas stand to benefit from findings emerging from corpus pragmatic research designs:

1. *Treatment of data.* Although the impetus for democratisation studies comes from observing actual language use, the treatment of data in many empirical studies is far from ideal. Studies that only rely on anecdotal evidence may be rare, but many studies are still based on fairly small, often opportunistic datasets, which are unlikely to serve as a solid basis for generalisations, especially if the method of analysis lacks systematicity.

---

1 Even studies that essentially follow a synchronic research design typically make use of some baseline data to argue that the empirical findings are indicative of change of some nature (e.g., Landert 2014)
2. **Limited temporal focus.** Most studies exclusively focus on the latter part of the 20th century, thus ignoring any prior changes whose description would be necessary in the construction of a more comprehensive description of the trend.\(^2\)

3. **Emphasis on general corpora.** Previous studies often rely on general corpora for drawing generalisations, ignoring variation at the level of registers and sub-registers (Biber and Gray 2013), whose role in the democratisation of discourse is often crucial.

4. **Terminological variation.** Although analyses invoking democratisation as an explanatory factor have strong intuitive appeal, the term, as well as related concepts like *conversationalisation*, *colloquialisation* and *informalisation*, has not been sufficiently well theorised with respect to the actual mechanisms under which it operates in actual language use.

5. **Overreliance on frequency information.** Most previous corpus-linguistic studies focus on frequency changes at the expense of detailed pragmatic analysis of the contextual features activated in different communicative situations, which would be crucial for interpreting the meaning of utterances and assessing the role of speaker agency and social factors in macro-level processes of linguistic change.

To address and engage with these (and other) issues, this special issue showcases six original studies on democratisation by an international team of authors. Before introducing the studies and how they contribute to the corpus-pragmatic perspective on democratisation, we discuss the different meanings of *democratisation*.

---

\(^2\) With some notable exceptions, such as Biber & Finegan (1989) and Smitterberg (2021).
2. Meanings of democratisation

Democratisation is a concept that is employed both in everyday language and in academic discussions. As a result, it encompasses different layers and shades of meaning, which are often left unspecified in actual usage situations. It is therefore instructive to distinguish between three common meanings of democratisation:

i. The non-technical, compositional sense of the term (*democratic* [adj.] + *ation* [suffix]), denoting the process of something becoming more democratic and assuming that it is clear what *democratic* means in the context in which the word is used;

ii. The highly technical usage in social science disciplines like political science, political philosophy, or comparative politics, which focus heavily on the societal dimension of democratisation and engage thoroughly with the different senses of *democracy/democratic* (e.g., Grugel 2001) but place less weight on language use (or even overlooks its potentially constitutive role);

iii. The technical sense in such areas of linguistics as (critical) discourse analysis and corpus linguistics, where the term is used to explore the interrelationship of language change and social change (e.g., Leech et al. 2009; Farrelly and Seoane 2012; Culpeper and Nevala 2012).

Of these three senses, the third one is the most central to our concerns, but it is clearly not independent of the other two.

With the help of corpora, it is easy to find attestations of the general, non-technical sense (i.) in contemporary journalistic discourse. Example (1), taken from the *Corpus of

---

3 Baker and McEnery (this volume) use the term *generic democratisation* to refer to this sense.
Contemporary American English (COCA; Davies 2008), features a talk-show host interviewing a businessman about his views on wireless technology in 1999.

(1) **Neil Cavuto:** You know, Sam Ginn, one of the areas that you’ve raised with me in the past is the fact that there’s been to paraphrase I think some industry experts, a great democratisation of wireless technology that it used to be for business, but now it's for just about everybody else and that that has changed the playing field markedly. You agree?

**Sam Ginn:** I think that’s right Neil.

(COCA, Spoken, 1999)

Here, Neil Cavuto’s use of the term is intended to informally refer to what he takes to be a consensus opinion among industry experts: that wireless technology will become increasingly accessible across society. Nearly a quarter of a century on, we can observe that although Cavuto hedges the statement about the fact that … there’s been a great democratisation with to paraphrase I think some industry experts, the at-the-time hypothetical evolution has been confirmed by recent history, leading to a true “digital media revolution” (Mair, this issue).

We can also use data from COCA to establish that, overall, example (1) does not represent the most typical usage of democratisation. Instead, Figure 1 shows that democratisation is clearly associated with the ACADEMIC section of the corpus, as indicated by its normalised frequency of 18.4 pmw, which is ten times higher than in MAGAZINES. The status of democratisation as an “academic” word is further underlined by the fact that out of the 3,071 instances in the COCA, 513 contain a postmodifying prepositional phrase headed by of (e.g., democratisation of the political process, democratisation of culture and government, etc.), which is commonly taken to be a token of academic style (Biber, Grieve and Iberri-Shea 2009).
Regarding the technical sense (ii.), Coppedge (2012: 11) laments the difficulties of studying democratisation, which include both reaching an agreement on what is meant by democracy, and how it is best measured:

Democracy is a contested concept because nearly everyone values the label, but there are different reasonable and legitimate, yet incompatible, criteria for judging whether the label is deserved.

Noting that there are over 600 definitions for democracy in the literature, he goes on to criticise the use of “checklists” as indicators of democratisation and argues in favour of “thicker” descriptions that are sensitive to the particularities of different contexts.

Perhaps surprisingly, these concerns about the specific denotations of democracy/democratisation are not necessarily shared by linguists and pragmaticians. Instead, when discussing democratisation as a discursive-pragmatic process (iii.), the term is applied in a general sense to hypothesise that (a) if a process of linguistic/discursive change can be identified to have taken place between points A and B in time, and (b) if the period between A and B also subsumes some kind of societal change, then these two processes may be connected. Some writers use adjectival premodifiers to highlight the fact that this sense of the term refers primarily to a linguistic process; thus, the terms discursive democratisation (Farrelly and
Seoane 2012) and *linguistic democratisation* are also encountered in the literature. It is probably fair to say that the theorisation of democracy and processes of social change is less nuanced in comparison, and understandably so, given that it would be difficult to postulate how minute details of organisation and governance of individual societies would influence the use or non-use of specific linguistic constructions, such as semi-modal or progressives, and to disentangle their possible effects from other determinants, such as technological developments or register variation. Accordingly, the literature on discursive democratisation typically refers to broad patterns of societal development as factors that are potentially responsible for linguistic change.

At the same time, that such a link should exist between societal change and linguistic change presupposes that the semantics of the linguistic features undergoing change are compatible with the proposed direction of societal change, and the theorisation of this aspect has sometimes been taken quite far based on limited evidence. For example, using the Google Books corpora as her material, Greenfield (2013: 1729) argued that the increasing frequency of words related to an individual’s freedom, wants and needs (e.g., *choose, get, self*) and the decreasing frequency of words pertaining to the needs of the community and the individual’s obligations in it (e.g., *duty, give, obedience*) are indicative of a more general shift from a rural to urban society and of people’s increased individuality at the cost of communal values. However, connecting frequency changes of individual words with macro-level societal developments can be problematic regardless of the size of the corpora used in the study. One of the problems concerns polysemy: words like *give* and *get*, for example, are highly

---

4 Conversely, the term *societal democratisation* is used by some writers to refer to the “institutional” (as opposed to linguistic) sense of the term.

5 These are probably best explored through case studies like Baker and McEnery (this issue) and Tyrkkö et al. (this issue).
polysemous, and they can plausibly be connected to psychological traits, such as a sense of communality and individualism, only in certain contexts of use (Hilpert 2020). Another challenge concerns the limitations related to the representativeness of linguistic corpora. As Hilpert (2020: 7) points out, if corpus data are taken at face value, the collocational profile of screwdriver would suggest that screwdrivers are dangerous weapons that are regularly used in violent crimes and burglaries, as evidenced by high-ranking collocates such as armed, stabbed, threatened, attacked, and forced. Indeed, if one fails to acknowledge the fact that the most mundane (and frequent) uses of screwdrivers are not newsworthy enough to be discussed in the texts included in the News on the Web corpus (which Hilpert used in his example), one can easily arrive at an analysis that may seem plausible from the perspective of corpus frequencies but which makes little sense from a societal perspective.

Another well-known issue with studies searching for correlations between societal change and linguistic change as manifested in corpus data (e.g., Leech et al. 2009) is that the method only tells one side of the story, as it does not directly engage with the role of language as an active agent of change through its potential to construct categories and impose specific world-views on the intended addressees. The constitutive role of language and discourse has been a central tenet in socio-cultural linguistics, where identity categories are seen as discursively constructed products that emerge in interaction (Bucholtz and Hall 2010). One example of such a discursively constructed category is a national identity, which is “produced, reproduced, transformed and destructed” through language and other semiotic systems (De Cillia et al. 1999). This perspective links up with Fairclough’s (1992) conceptualisation of democratisation as the process of “removal of inequalities and asymmetries in the discursive and linguistic rights, obligations and prestige of groups of people”.
The question of how exactly discourse shapes political/social democratisation was identified by Farrelly and Seoane (2012: 399) as worthy of further empirical work in historical linguistics more broadly. A significant amount of work in linguistics has been devoted to identifying other processes which are (either closely or loosely) linked to democratisation: these include *colloquialisation* (Hundt and Mair 1999; Mair 2006; Leech et al. 2009), *conversationalisation* (Fairclough and Mauranen 1997), *informalisation* (Fairclough 1992), *Americanisation* (Gonçalves et al. 2018), *technologisation* (Fairclough 1995), *commodification* (Heller 2010), *popularisation* (Biber and Gray 2012), *tabloidisation* (Holly 2008), *stylistic drift* (Biber and Finegan 1989), and *stylistic levelling* (Schützler 2020).6

In this introduction, our aim is not to provide a comprehensive overview of the relevant terms and their different shades of meaning. Useful overviews exist in previous literature, which discuss how the notion has been conceptualised in previous work on the history of English (Farrelly and Seoane 2012), how it relates to other sociocultural processes (Culpeper and Nevala 2012), how it links with gender-neutrality (Loureiro-Porto and Hiltunen 2020) and in general how it can be empirically studied (Hiltunen and Loureiro-Porto 2020). Having said that, the plethora of terms might also be indicative of a certain looseness of terminology and an excessive tolerance of terminological variation. While such a situation is probably common in pluralistic and divergent disciplines such as linguistics – and indeed similar concerns have been expressed for other subfields as well (e.g., study of formulaicity, see Wray and Perkins 2000) – in the worst case it might make comparing the results of individual studies difficult. For this reason, it is essential that individual studies always specify how the notion of democratisation is employed in the context of specific study designs (which is what the authors

---

6 In this context, we should also mention terms for processes that work in the ‘opposite’ direction, like *anti-colloquialization* and *densification*, which is often investigated simultaneously with *colloquialization* (e.g., Smitterberg 2021).
in this special issue do). In this sense, the critical evaluation of colloquialisation in Mair (this issue) is a particularly welcome contribution to the current understanding of this phenomenon.

3. Towards a corpus-pragmatic perspective on democratisation

The present collection of articles elucidates the multifaceted phenomenon of discursive democratisation from different angles. Related to the theory of democratisation and more specifically colloquialisation, the issue opens with Christian Mair’s article, where he takes stock of the last twenty-five years (or so) of colloquialisation research. The usefulness of the notion of colloquialisation is amply evidenced by the large volume of corpus-based studies devoted to it, which can be attributed to the fact that the notion can be easily operationalised in corpus studies with general corpora, provided that their structure incorporates different modes (spoken–written) and levels of formality in an appropriate way. At the same time, such studies may run into difficulties with the subtle differences between individual contexts of writing and speech, as well as with the inherent variability of language use. Related specifically to these, Mair identifies and outlines two elusive issues that have hitherto not been fully resolved: the difficulty of establishing a convincing baseline against which a purported colloquialisation trend could be identified, and the problem of defining the envelope of variation for individual features in L2/multilingual settings. But despite these challenges, Mair is highly optimistic about the continuing usefulness of colloquialisation for making sense of increasingly pluralistic media landscape and recognising discursive threats to democracy and democratisation in an institutional sense.

In terms of research design, the articles by both Elena Seoane and Lucía Loureiro-Porto and Turo Hiltunen and Turo Vartiainen build substantially on earlier scholarship on colloquialisation but incorporate important context-dependent adjustments, brought about by a critical review of theoretical concepts and the use of data as a source of empirical evidence.
Seoane and Loureiro-Porto focus on discursive democratisation from the perspective of the pragmatic negotiation of power relations by investigating three pragmatic variables: core vs. semi-modals, contracted vs. non-contracted forms, and passive vs. active verb constructions. The variables are carefully selected to provide evidence of the gradual democratisation of English as used in New York Times editorials from different angles. The quantitative results of Seoane and Loureiro-Porto’s case studies offer clear evidence of a general democratisation trend, and their carefully contextualised pragmatic analyses provide even more information about the relationship between sociohistorical events and language use. For instance, although the frequency of the core modal must is on the decrease in the latter part of the twentieth century in their data (in accordance with the general democratisation of journalistic discourse), there are conspicuous peaks in its usage that coincide with World War I and World War II (1917, 1940, 1942) and the time when the public debate about the Vietnam War was reaching its peak (1970). The authors regard these periods as “counter-episodes in the democratisation of English” brought about by a need to use direct and unambiguous discourse in wartime. The detailed pragmatic analyses of contracted forms and passive constructions, on the other hand, show that not only do their frequency changes support the hypothesis of discursive democratisation, but the functional changes they are indicative of can also be explained by democratisation.

The focus of the study by Hiltunen and Vartiainen is on parliamentary discourse, which is a register at the intersection of spoken, written, and written-to-be-spoken language (see Hiltunen et al. 2020). As such, it is conveniently positioned for the exploration of questions related to the colloquialisation of public discourse. Hiltunen and Vartiainen investigate whether differences in the pace of colloquialisation can be identified between the House of Commons and the House of Lords, following the recommendation by Biber and Gray (2013) that linguistic variation should be studied at the level of maximally specific sub-registers. The relevance of
this perspective becomes evident in the authors’ analysis, which shows that shifts in discourse styles do not proceed at the same rate in the two Houses of Parliament, which Hiltunen and Vartiainen propose is largely due to differences in the editorial conventions of the Official Reports that reproduce the speeches delivered in the two Houses. In addition to a corpus-pragmatic study of the relevance of sub-registers in linguistic analysis, the authors provide a more qualitatively oriented case study of a rhetorical pattern that has not been discussed in previous research. In this pattern, the speakers use a third-person coreferential construction with *think* to indicate their disagreement with the referent of the subject pronoun (e.g., *he thinks he has answered the debate but he has not*). This case study not only delves deeper into the pragmatics of the parliamentary debates but also reminds us of the importance of critically evaluating the design of frequency-based investigations: by moving from specific usage patterns to more abstract categories (e.g., lemmas), one runs the risk of glossing over important functional differences.

The remaining three studies focus on questions of democratisation in increasingly specific contexts of political discourse, as they investigate group construction processes and discursive representations of different political actors in political speeches and newspaper articles. Jukka Tyrkkö, Sophie Raineri, Jenni Räikkönen, Alžběta Budirská, Mai Nabawy and Amanda Silfver focus on political speeches delivered by activists, who are outside the mainstream party-political arena and campaign for causes such as civil rights and female suffrage. Activist speeches have remained an under-researched area in discourse-pragmatic studies despite their crucial role in democratisation and in causing societal and political change. By combining quantitative and qualitative methods, Tyrkkö et al. focus on diachronic changes in the rhetorical use of personal pronouns in a corpus consisting of 120 American activist speeches and on the ways in which the pronouns are used to construct ingroups and outgroups. While the diachronic changes in the frequencies of the pronouns follow the same general trend
as has been observed in previous studies focusing on mainstream political speeches, constructing ingroup/outgroup dichotomies also involve atypical uses, such as using *they* to refer to the ingroup. Furthermore, the speeches show a trend toward more generic engagement, as there is a decline in outgroup references, which suggests that activist speakers decreasingly focused on the identities of “us and them” and more on activating the general public.

Continuing with the theme of activism, **Birte Bös** examines features of discursive democratisation in a specialised corpus comprised of articles from three suffrage newspapers from 1912–1913. Suffrage periodicals played a key role in women’s fight for universal suffrage and increased democracy. The study combines corpus-linguistic methods, such as keyword analysis and concordances, with a close-up qualitative analysis. The data are examined from a discourse-pragmatic perspective, and Bös particularly focuses on group construction processes indexed by features such as person references and actions connected to different social actors as well as linguistic features of immediacy. Bös concludes that a frequent use of micro-level linguistic features such as the first- and second person pronouns and *be* as a main verb as well as more macro-level features such as directives were used to create a sense of immediacy and closeness in order to activate the readers for the suffrage movement. These features of linguistic immediacy can also be seen as contributing to democratisation, as they level out hierarchical differences and make texts more accessible and reader-friendly. However, as noted by Bös, these were counterbalanced by linguistic features of distance such as indeterminate pronouns and passive constructions, which suggests that the suffrage press also adopted some of the practices of professional news writing typically refraining from colloquialisation.

**Helen Baker** and **Tony McEnery** analyse newspaper articles reporting on the Liberal politician William Gladstone’s Midlothian campaign (1879–1880), which has been argued to be the first British electoral campaign with a direct appeal to the public through hustings and the print media. The Midlothian campaign offered Gladstone the means to communicate
directly with the public and to create a media event that would make him more visible in the print press and familiar to the people. Employing Social Actor Analysis (van Leeuwen 2008), the paper examines the choices made by the newspapers in reporting about the campaign and, in particular, about the Gladstone family. A collocation analysis, which shows the prominence given to the members of the Gladstone family in the newspaper reports, suggests that Gladstone was represented as a family man, with whom the electorate of that time could easily identify. Furthermore, Baker and McEnery highlight the differences in reporting about the members of the Gladstone family. For instance, William Gladstone was accorded honorifics, while his wife, Catherine Gladstone, was merely established as being married to Gladstone through the use of Mrs. The analysis complements the previous research done on the Midlothian campaign, showing that Gladstone understood well the power of image and the media in strengthening his support amongst the public.

4. Conclusion

To conclude this introduction, we return to the challenges that we discussed at the beginning of the chapter: the treatment of data, limited temporal focus, emphasis on general corpora, terminological variation, and an overreliance on frequency information. We would argue that the articles in this special issue make substantial headway in meeting these challenges. First, the studies make use of large corpora and text collections in ways that permit the data to be studied from both qualitative and quantitative perspectives. This perspective to linguistic analysis is precisely what underlies the corpus-pragmatic approach as envisioned in Rühlemann and Aijmer (2015). Indeed, instead of just observing changes in the frequency of specific linguistic features and patterns, the authors engage in careful contextual analyses of their data, which not only makes the value of their quantitative results more apparent but also allows the authors to connect their work with socio-historical developments that have unfolded in the past.
two centuries, and to democratisation in particular. The socio-historical interpretations offered by the authors are further bolstered by the long time span covered in the studies: by examining linguistic developments over long periods, the authors are able to identify specific points of interest in their material and to mitigate the potential effect of random fluctuations in the data.

Second, instead of studying linguistic developments in general corpora, all the authors in this special issue show great awareness of the importance of the sub-register. While some of the studies make use of corpus resources that are generally available for research purposes (e.g., the Hansard Corpus or A Corpus of Historical American English), others have compiled specialized corpora in order to study pragmatic developments that are mediated by specific sub-genres (political speeches, newspaper articles and editorials). Although corpus compilation is a labour-intensive task, we argue that it is often necessary if one wishes to reach reliable conclusions about how grammatical and discourse-pragmatic changes unfold over time. Thanks to the increased availability of historical public texts, it is now possible to engage in corpus-pragmatic analyses of a variety of registers. Furthermore, while the use electronic archives often raises new questions regarding the best practices of using them as corpora (e.g., Hiltunen 2021), future studies will undoubtedly increase our understanding of the specific pathways of linguistic change and the relevance of the sub-register.

From a methodological perspective, the articles raise questions that in our opinion are relevant to any corpus-pragmatic study. Mair’s observations of varying baselines and multilingual contexts are of course central here, and the latter point in particular links with a larger issue: communication (both online and offline) in today’s increasingly global and diverse communities inevitably gives rise to a “sociolinguistic superdiversity” (e.g., Creese and Blackledge 2018) with a new set of issues for work on democratisation. Theoretical and empirical advances in this respect would seem to depend both on cross-linguistic work investigating democratisation processes in different languages (e.g., Pons Bordería 2014 on
Spanish; Skärlund 2018 on Swedish) and on methods that are able to incorporate language contact phenomena in the study design.

Another methodological question that in our opinion would merit further consideration is the multifunctionality of linguistic constructions indexing colloquial style: as demonstrated in Hiltunen and Vartiainen’s contribution, forms that on the surface appear similar may on closer scrutiny turn out to be very different in their function, and this issue should be taken into account in the analysis. While a functional analysis of every single form under study may be unrealistic especially with large corpora, a recognition of the multifunctionality of linguistic constructions underlines the importance of a careful contextual analysis as one component of a historical corpus-pragmatic study.

The studies included in this special issue reflect the wide scope and pluralism of current work in historical corpus pragmatics, which in turn is a logical consequence of the richness of the phenomena that fall under the purview of this area of study. There is a lot to say about this pluralism, and the findings of individual contributions demonstrate that the field clearly benefits from interdisciplinary collaboration (e.g., combining history and linguistics) as well the exploration of new materials and systematic attempts to improve the accuracy of earlier research. At the same, the divergent nature of the field of pragmatics is an obvious challenge to any attempts to establish a coherent research agenda that could be pursued collaboratively and collectively. Looking ahead, one desideratum for future work would be the identification of best practices for achieving this, not only in the context of democratisation but also more generally. In other words, while it is difficult to disagree with the necessity of multiple perspectives in historical corpus-pragmatic work, exactly how this can be accomplished is, and may remain, a question that needs to be answered on a case-by-case basis.
References


