ELSEVIER

Democratization of Englishes: synchronic and diachronic approaches

Turo Hiltunen [a,*], Lucía Loureiro-Porto [b,**]

[a] The Department of Languages, University of Helsinki, PO Box 24, 00014, Finland
[b] Dpt. Filologia Espanyola, Moderna i Clàssica, Universitat de les Illes Balears, Edifici Ramon Llull, Cra. Valldemossa Km. 7,5, 07122, Palma de Mallorca, Spain

ARTICLE INFO

ABSTRACT

The term *democratization* has been used in recent linguistic research to describe how specific linguistic changes can be linked to changes in sociocultural norms. This broad definition, however, does not fully capture the essence of this phenomenon or explain how it differs from other processes of language change. Other key issues in this area of research include what the cause-effect relationship is between linguistic change and social change, and how empirical corpus linguistic studies can contribute to current knowledge. In this opening contribution to the special issue *New perspectives on democratization: Evidence from English(es)*, we address some of these key issues by reviewing previous synchronic and diachronic work studies on democratization in different varieties of English, and introduce new studies that take evidence from different linguistic corpora. By placing the linguistic changes into their specific socio-historical contexts, these studies yield interesting results, showing that variationist linguistic methodology may significantly contribute to disentangling the complex relationship between language change and social and societal changes.

1. Introduction

A growing volume of work in English linguistics addresses issues related to the relationship between language and society, using the term *democratization* as an explanatory parameter. Originating in Critical Discourse Analysis (e.g. Fairclough, 1992), the term *democratization* (of discourse) refers to how socio-cultural changes related to "changing norms in personal relations" (Leech et al., 2009: 259) are linked to linguistic changes which can be observed in different English-speaking territories from the 19th century onwards. Work in this area commonly departs from some version of the hypothesis that as societies become more democratic (at least in appearance), expressions of power inequalities in public discourse become less frequent (see e.g. Smith, 2003; Baker, 2010). Many recent and ongoing changes in English have accordingly been linked to democratization, including the decline in the use of deontic modals (Myhill, 1995), titular nouns (Leech et al., 2009), and gendered pronouns and nouns (e.g. Baker, 2010; Paterson, 2014).

---

* Corresponding author.
** Corresponding author.
  E-mail addresses: turo.hiltunen@helsinki.fi (T. Hiltunen), lucia.loureiro@uib.es (L. Loureiro-Porto).

However, "democratization" understood in this sense is a complex phenomenon which defies simple analysis. First, discursive democratization is mediated through a variety of situational, cultural and political factors, which makes generalizations and cross-study comparisons difficult. Secondly, language does not exist independently of social and societal categories and does not merely represent or reflect them; rather, using language is a social practice in itself, and one that interacts with social categories and influences and constructs them (Cameron, 1990; Baker, 2010). Finally, establishing links between linguistic observations and sociocultural factors needs to be done with care to avoid explanations that are simplistic or even misleading, and this requires close attention to and a critical attitude towards data (e.g. Oakes and Farrow, 2007; Baker, 2010; Koplenig, 2017).

Many of these issues are addressed in this special issue. Taking corpus data as the point of departure, the contributions investigate language use as part of a larger sociocultural system that changes in interaction with different tendencies in history, with the aim of shedding new light on the process of linguistic and discursive democratization.

This introduction begins with an overview of the concept of democratization and related concepts (Section 2). It then moves on to discussing some current debates surrounding the issue of democratization that have emerged from previous work (Section 3), with specific reference to questions of definition, operationalization and methodology, which are also taken up in the contributions. Finally, Section 4 gives an overview of the papers and assesses their contribution to the theoretical debates on democratization of discourse, concluding with some thoughts on potentially fruitful areas of further research.

2. What is meant by democratization?

The most basic sense of the term democratization is "making something more democratic in organization or character" (OED) and "the process of becoming a democracy" (*Concise Encyclopedia of Democracy*, Staff of the Congressional Quarterly, 2013), and it is most often applied to societies (see also Kotze and van Rooy, this issue). Perhaps the best-known discussion of democratization in this sense is Samuel Huntington's *The Third Wave: Democratization in the Late Twentieth Century* (Huntington, 1991), where he puts forward the idea of democratization occurring in waves, which are followed by reverse waves of authoritarianism (Grugel, 2001: 32–37): the expansion of liberal democracies in the first wave (1800–1930) was cut short by the rise of communism and fascism, and the second wave, beginning in 1945, was punctuated by military coups and emerging dictatorships in Latin America. The collapse of authoritarian rule in Portugal, Spain and Greece in the mid-1970s marks the beginning of the third wave, which has recently been reversed by what Diamond describes "a mild and protracted democratic recession" (Diamond, 2015: 144), evident in the declining number of democratic regimes and the resurgence of authoritarian and populist leaders across the globe. This is largely confirmed in the recent report by the *International Institute for Democracy and Electoral Assistance*, which concludes that "democracy continues to expand its global reach, while experiencing a significant decline in quality across the board" (2019: vi).

Democratization in this sense is partially related to the sense in which the term is typically used in linguistics, although there are clear differences. The popularization of the term in linguistics and discourse studies can be attributed to the work of Norman Fairclough (e.g Fairclough, 1989, 1992), who defines democratization as "the removal of inequalities and asymmetries in the discursive and linguistic rights, obligations and prestige of groups of people" (Fairclough, 1992: 201). For Fairclough, discourse has been significantly influenced by the process of democratization, which is manifest in the increased acceptance of informality and non-standard varieties in public and institutional discourse, as well as the reduction of overt markers of power.

In their overview of previous work, Farrelly and Seoane (2012) describe democratization as a discourse-pragmatic process and identify three main areas that have been the focus of previous research: the phasing out of overt markers of power asymmetry ("democratization proper"), a shift to a more speechlike style (*colloquialization*) and a tendency towards informality (*informalization*) (Farrelly and Seoane, 2012: 393). Democratization proper is manifest in the *decrease* of various linguistic features marking social hierarchy and gender (titular nouns, gendered pronouns and nouns, modals expressing strong obligation) and the concomitant *increase* of more egalitarian expressions that emphasize solidarity and minimize the imposition on the addressee (e.g. gender-neutral pronouns/nouns and modals expressing median obligation).

Colloquialization and informalization are closely related, and sometimes used interchangeably. As established in genre and register studies, spoken English is characterised by a relatively high frequency of such features as contractions, progressives, personal pronouns and discourse particles,[1] and the increase of these features in written registers can be interpreted as a consequence of a process of colloquialization (e.g. Mair, 1997; Leech, 2004; Mair, 2009).[2] Informalization relates more closely to changes towards the increasing reader-friendliness and accessibility of traditionally formal registers like newspaper writing and scientific prose (Farrelly and Seoane, 2012: 395–396). It is also closely related to what Fairclough (1989) and Landert (2014) have termed (synthetic) *personalization* of mass media communication: the foregrounding of individuals to simulate face-to-face interaction and create involvement, and to *conversationalization*, understood as the simulation of conversation in public discourse to attract wider audiences (Fairclough, 2003; Smith, this issue).[3]

---

[1] For a more comprehensive list of features, see e.g. Leech (2004: 75).

[2] There is some variation in the use of terminology: Biber and Finegan (1989) use the term *drift* to refer to a change observed in written genres towards more "oral" styles between the 17th and the 19th centuries. The increase of colloquial forms is referred to as *popularization* in Biber and Gray (2012), a tendency which is contrasted with *economy*, namely the aim to present more information in a concise way, characteristic of specialized subregisters of academic prose.

[3] A related term is the *tabloidization* of broadsheet journalism, which refers not only to changes in journalistic content (e.g. the increase of news about entertainment and celebrities) but also to the way in which language is used (e.g. the reduction of length of words and sentences and complexity of language) (Conboy, 2010: 130–133). See also Lefkowitz (2018).

The fact that the study of linguistic and discursive democratization is an active area of research in English linguistics is not surprising, given that accumulating evidence increasingly supports the view that many recent changes in English are not exclusively due to factors internal to the language system. Instead, they can be attributed to general sociocultural changes towards informality in manners and a growing emphasis on equality, which is manifest in language use as a preference for informal and "democratic" expressions, which can be observed in corpora (Mair, 1997: 204). The interplay between the external and the internal motivations for recent language change is far from being conclusively described and, for that reason, discerning whether the change originates in sociocultural changes or is caused by intralinguistic factors is one of the key issues discussed in research on democratization, as seen in the next section.

3. Key issues and current debates

As is obvious from the previous section, democratization is far from being a monolithic phenomenon, with specific characteristics that distinguish it from other, related processes. Moreover, it must not be overlooked that there are a series of issues that nurture the debate regarding its study and characterization. In the paragraphs that follow, we offer a brief discussion on three of them, namely (i) the directionality of democratization of language/discourse, (ii) the contribution of corpora and variationist linguistics methodology, and (iii) identification of features associated with democratization which have not been described earlier.

Does language reflect society, or does society reflect language? The question of the direction of the relationship in linguistic change and sociocultural change is an important one in the study of linguistic and discursive democratization. Implicit in most analyses is the idea that societal and cultural changes have an impact on how language is used, and can often be observed as changes in the text frequencies of features, as previously discussed (see e.g. Leech and Fallon, 1992; Oakes and Farrow, 2007; Leech et al., 2009). One illustration of this concept is the decrease of gendered pronouns observed in the Brown family of corpora, which has been linked to societies becoming more democratic, and as such less likely to show gender-based bias (Baker, 2010: 69). It is of course often difficult to pinpoint exactly how this process works. The explanation of some cases appears fairly straightforward (and at the same time linguistically trivial according to Szmrecsanyi, 2015: 3): for example, the increased use of the word *computer* in the *Google Books Corpus* in the second part of the 20th century clearly corresponds to technological advances (Fruehwald, 2016: 45). In general, however, aspects of society and culture potentially affecting language use are complex and defy simple operationalization as linguistic variables. But as Culpeper and Nevala (2012: 365) put it, even though sociocultural processess "lack the apparent certainty of formal linguistic material", they too can be tackled with scholarly rigor. Incorporating sociocultural factors in the analysis requires careful appraisal of the contextual configurations of corpus texts combined with close reading (see Hiltunen et al., this issue; Smith, this issue), and these qualitative analyses are not easily compatible with the use of large corpora, yet this is clearly desirable to safeguard against simplistic interpretations (Jockers, 2013; Brezina, 2018).

On the other hand, many studies especially in Critical Discourse Analysis (CDA) reject the "language reflects society" model described above on the grounds that language is instrumental in constructing and maintaining the power structures in society (e.g. Fairclough, 2010). Similarly, in his book on the history of newspaper language, Conboy (e.g. 2010: 4–5) criticizes the idea that language could be separated from its sociocultural context of use, arguing instead that language should be seen as an integral part of the social, interacting with social categories and also constituting them. For Farrelly and Seoane (2012), democratization encompasses the dual relationship between language and sociocultural change, such that "people alter their language use in response to social change and people shape social change through their use of language" (2012: 392).

How do we get new evidence from the core processes of democratization? Another key issue for the study of democratization is the question of evidence: how can we get evidence for confirming or disproving specific hypotheses about democratization? To answer this question, the contributions to this special issue turn to linguistic corpora. This approach dovetails well with the "quantitative turn" in different areas of linguistics (e.g. Janda, 2013; Kortmann, 2018), driven by the proliferation of very large corpora (for an overview, see e.g. Hiltunen et al., 2017) enabling the use of techniques of quantitative corpus linguistics and exploratory data analysis. An example of quantitative analysis of very large corpora is, for instance, Gonçalves et al. (2018), which explores two large datasets, one based on tweets published between 2010 and 2016, and another one based on *Google Books*, with the aim of mapping the Americanization of English. However, the fact that the results of these big data analyses overlap to a large extent with those obtained using more traditional corpora – such as the Brown family of corpora and *ARCHER* (Mair, 2019: 340) – highlights the value of small and tidy corpora in the study of grammatical change in general. This being the case, the dictum "small is beautiful" (Hundt and Leech, 2012) (see also Loureiro-Porto, 2017) clearly applies to studies on linguistic and discursive democratization, and papers in this issue accordingly explore both carefully-curated corpora and large datasets, both separately and in combination with each other (see e.g. Palander-Collin and Nevala, this issue).

The advantages of corpora in linguistic analysis are numerous (for a review, see e.g. Biber, 2010) but here we would like to highlight only one: corpora enable the systematic analysis of natural language produced in different environments. This, as is well-known, has an important consequence for studies on linguistic variation: as corpora serve as representative (and in principle non-biased) samples of authentic language use, the observer can analyse the language objectively, without allowing any intra-personal characteristics to interfere with the scientific interpretation of data. This is a major convenience compared to discourse-analytical qualitative studies based on a small amount of data. Having said that, qualitative analysis is often needed for assessing the motivations and communicative intentions that give rise to the trends observed in corpora.

There are other issues associated with the use of corpora for studying democratization processes. In addition to the thorny issue of corpus representativeness in general,[4] many corpora do not provide specific information about discourse participants – their ages, their relation (hierarchical or egalitarian), their genders, education, etc. – and the specific effects of these extralinguistic factors cannot therefore be properly assessed.[5] It is precisely for this reason that triangulation of methods may become necessary when trying to determine whether the variation (or change) is explained as concomitant with social changes leading to a higher equality or not.

How can we identify less obvious patterns? Another question is the identification of linguistic features associated with democratization and other related processes. It could be argued that the main linguistic correlates of democratization are well-known. Previous work on democratization proper has focused on modal verbs (e.g. Myhill, 1995), titular nouns (e.g. Leech et al., 2009: 259–260), and gendered pronouns (e.g. Laitinen, 2007; Paterson, 2014), on which plenty of other studies are available. The range of features associated with colloquialization, informalization and popularization is even broader, including many of those features identified as marking involvement in Biber (1988) and associated with the spoken register in Biber et al. (1999). This approach is followed by several studies in this special issue (Kranich et al., this issue; Kotze and van Rooy, this issue; and Schützler, this issue, the latter focusing on the connective *though*, which has previously not been linked to democratization).

However, as pointed out by Rühlemann and Hilpert (2017: 105), restricting the attention to these features is problematic, as it runs the risk of missing potentially relevant features that have not been identified in previous studies. As an alternative, they suggest an approach where the point of departure is a corpus-driven investigation of the data with minimal assumptions (Tognini-Bonelli, 2001), which is then followed by a corpus-based investigation of those specific features which emerge as relevant in the data. Along with discovering features specific to different datasets, this approach has the advantage of highlighting less obvious features, whose relevance to democratization can be confirmed with the close analysis of examples in context. Four studies in this special issue use a combination of corpus-driven and corpus-based approaches: Hiltunen et al. (this issue) carry out a data-driven analysis of n-grams in the *Hansard Corpus* using hierarchical cluster analysis and principal component analysis, Palander-Collin and Nevala (this issue) use a keyword procedure to identify key "people words" in the ARCHER corpus and investigate their frequencies in large newspaper databases, Lundberg and Laitinen (this issue) explore frequency-ranked lists of unigrams and bigrams, and Smith (this issue) applies a keyword procedure to POS-tags to identify features for closer analysis in the BBC Radio 4 chat show *Desert Island Discs*.

4. Democratization of Englishes: synchronic and diachronic approaches

Approaching the democratization in various varieties of English from different angles contributes to addressing many of the key issues and current debates discussed in Section 3. With that purpose, the papers included in this special issue explore the notion of democratization from both a diachronic and a synchronic perspective, and relate linguistic democratization, as evidenced in different corpora, to historical changes leading to societal equality. Thus, a diachronic analysis of democratization is found in:

- Hiltunen et al. (who explore the British Hansard Corpus between 1870 and 1930),
- Palander-Collin and Nevala (who investigate both the British component of ARCHER and different "big data" corpora, as found in *Google Books*, *British Library Newspapers* and *The Economist*, for the period 1700s–1900s),
- Smith (who uses a specialized corpus based on a radio programme aired from 1960s onwards)
- Schützler (who compares the diachronic findings in ARCHER to synchronic data extracted from ICE),
- Kotze and van Rooy (who explore the British, Australian and South African Hansards from 1901 to 2015)

The synchronic analyses presented in this issue include:

- Schützler (on different varieties of English as found in ICE, namely BrE, AmE, CanE, IndE and NigE),
- Kranich et al. (who combine the analysis of GloWbE for different varieties with discourse completion tasks, DCTs, in BrE, AmE and IndE),
- Lundberg and Laitinen (who explore English as a Lingua Franca (ELF), as found in tweets produced by humans and by bots).

As is clear from the previous list, the varieties of English analysed in this issue range from inner-circle varieties (mainly BrE, but also AmE, AusE and CanE) to outer-circle (SAfE, IndE, NigE) and expanding circle ones (ELF). This richness in the number of varieties explored is complemented by a wide range of methodological techniques: some studies are corpus-based, while others are corpus-driven and based on different methods of quantitative corpus linguistic and statistical techniques including

---

[4] The discussion of the extent to which a selection of texts can represent a whole variety is one we will not enter here (see e.g. Sinclair, 2005; Evert, 2006).

[5] A related problem is inaccurate metadata about texts and writers, associated especially with the use of large text collections as corpora (Koplenig, 2017).

keyness and type-token ratio analysis, negative binomial regression models, clustering analysis and principal component analysis, among others.

In addition, these papers relate their findings to relevant socio-political changes, such as the "liberal" legislation that took place in 19th-century Britain (discussed in Palander-Collin and Nevala, this issue), changes in the recording of parliamentary records (Hiltunen et al., this issue), and the access of previously disenfranchised groups to the South African Parliament (Kotze and van Rooy, this issue). Thus, all papers address the issue of *pseudo-democratization* (Mair, 1997; Clayman and Heritage, 2002), that is, the phenomenon of democratization is viewed by a critical eye, so as to bring to the fore (i) the idea that what appears a more democratic discourse does not necessarily entail a more egalitarian society; (ii) the fact that the Western-centric concept of linguistic democratization may have different interpretations in outer-circle varieties of English (in line with (Hansen, 2018: 244)); and (iii) the intersect between democratization and related phenomena such as colloquialization and conversationalization. The main findings in these three regards are summarized in the paragraphs that follow.

The study of the democratization of the language used in the British Parliament found in Hiltunen et al.'s article (this issue) is possible thanks to the existence of the Hansard Corpus (Alexander and Davies, 2016), in many ways "a corpus linguist's dream" (Mollin, 2007: 187). By analysing n-grams through a pattern-driven approach (Kopaczyk and Tyrkkö, 2018), the authors unveil many changes related to colloquialization (the second main type of democratization, according to Farrelly and Seoane (2012)) observable in the language used in the House of Commons from 1870 to 1930. This period is particularly interesting because the historical decision to introduce the Official Report in 1908 explains the transition from a transcription expressed in third-person summaries of the debates to first-person reporting, which entails a series of stylistic changes. In fact, this socio-historical fact explains the patterns that emerge from the analysis of 3-grams in the Hansard Corpus: clustering and principal component analysis of the patterns reveals that after 1909 the variety used in Parliament is considerably more colloquial than before that date, which constitutes a major stylistic shift related to democratization that would be difficult to identify and describe via other methods. In addition to identifying this major stylistic change, the study focuses on the qualitative analysis of two 3-grams, namely *is going to*, clearly associated with the near-verbatim report introduced with the Official Report, and *I think it*, a hedging device or an epistemic marker used to signal the speaker's personal opinion. The individual frequencies of these two 3-grams rise dramatically after 1909 but fall again in the 1920s, which is interpreted by the authors as a consequence of the increasing editorial control of *Hansard*. Thus, the combination of a mixed-methods approach, the analysis of individual features and a close reading of the historical events yields very relevant results for the interpretation of colloquialization, understood as a sub-type of democratization.

Another process closely related to democratization is conversationalization, and the interrelationship of these processes is studied by Smith (this issue) in a diachronic corpus representing the BBC radio 4 show *Desert Island Discs* (DID). To explore whether a field discourse can become more conversational and less democratic and vice versa, Smith carries out a detailed analysis of grammatical features (identified using a key part-of-speech tags method), complementing previous pragmatic analyses of the language of DID by Jucker and Landert (2015) that focused on turn-taking and turn length. Smith considers two periods with the first host (the 1960s and early 1980s) and two with the second long-serving host (the late 1980s and early 2000s), which enables him to identify some differences between DID and conversation, and also between conversationalization and democratization. Although these two processes converge in the use of stance and discourse markers, hesitations, first-person singular pronouns and generic *you*, they also diverge regarding the use of continuers and deictic *you*. In Smith's view, a host who withholds continuers (e.g. *mhm*) might seem less cooperative at the conversational level, but more democratic towards the audience. Likewise, the use of deictic *you* is only infrequently used by guests to address the host, which, while reducing the conversational flavour of the discoure, invites the audience to feel more included in the talk and can therefore be interpreted as more democratic. Overall, the analysis of the DID corpus reveals that even though conversationalization and democratization overlap to a large extent, they do not always converge.

Another paper that explores the interplay between colloquialization and democratization is Schützler's (this issue). Focusing on both inner-circle and outer-circle varieties, he proposes a taxonomy of democratization that includes "explicit democratization" (referring to linguistic changes that make direct reference to changes in social reality) and "textual/ stylistic democratization", which subsumes colloquialization and informalization, both of which involve what Schützler calls "stylistic levelling". "Stylistic levelling" is considered to be behind the increasing frequency of the conjunct *though*, which constitutes a less direct way of expressing disagreement than other markers such as *but* or *however* (Conrad, 2004: 72). Thus, the study of the conjunct *though* in several text-types in ARCHER (BrE and AmE) and ICE (BrE, CanE, NigE and IndE) via negative binomial regression models shows clear diachronic evidence of stylistic levelling/colloquialization/ democratization in BrE and, most clearly, AmE, though the inspection of synchronic varieties is less informative in this respect. While the two outer-circle varieties (NigE and IndE) exhibit a higher frequency of *though* in writing, this should, in Schützler's view, not be considered evidence of the higher colloquialization of those varieties, but should rather be interpreted as a result of the early stage in their development as postcolonial varieties in terms of Schneider's (2007) model. According to this model, diversification along socio-stylistic lines is possible at stage 5, which neither NigE nor IndE have so far reached. The paper rounds off with some suggestions regarding the interpretation of its results: (i) disentangling socially-motivated changes from mere grammatical ones, such as *grammaticalization*, may be hard, (ii) for a full approach to democratization the study of a single feature must be taken with some caution, and (iii) in the context of World Englishes, language contact between L1 and L2 must not be overlooked.

The diachronic democratization of BrE is also the subject of Palander-Collin and Nevala's paper (this issue). In order to address some of the same issues identified by Schützler, they combine corpus linguistics and socio-pragmatic methods, with the aim of triangulating the results. Thus, after a keyword analysis of people words in ARCHER, they explore three "big data" archives (*Google Books*, *British Library Newspapers* and *The Economist*). Interestingly enough, both kinds of data exhibit very similar general trends: the 19th century constitutes a turning point in the use of people words in British English. While before that date, keywords refer to individual persons and to their hierarchical order (e.g. *king*, *gentleman*, *madam*, etc.), the 20th-century data produce keywords referring to collective groups of people (e.g. *patients*, *workers*, *leaders*). This change in person reference correlates with important societal changes that took place in the 19th century, such as the enacting of "liberal" laws that contributed to protecting the most vulnerable members of society, and the improved position of individuals in the labour market. The study is complemented by a detailed analysis of two keywords, *gentleman* and *workers* in ARCHER, reaffirming the idea that small, carefully-compiled corpora are as useful for pattern identification as big data ones.

Data triangulation is also one of the pillars of Kranich et al. (this issue), which aims at unveiling the ways in which a society's decreasing attention to social hierarchies plays a role in the changing frequencies of markers of modality – the decreasing frequency of core modals and a corresponding increase of less face-threatening semi-modals is considered a paradigmatic example of alleged democratization in many varieties of English (Leech et al., 2009). Following a comparative analysis of the frequencies of core modals in BrE, AmE and IndE (as found in *GloWbE*), the authors implement a novel analysis of the use of modals and semi-modals in requests based on variational pragmatics (Barron and Schneider, 2008), by distributing Discourse Completion Tasks (DCTs) to British, American and Indian speakers of different ages. Interestingly enough, the hypothesis that Indian English speakers would be most influenced by hierarchical power structures (based on the overall frequencies of the core modals in GloWbE) is not confirmed by the DCTs. Rather, the IndE speakers exhibited the highest frequency of use of direct requests, while the BrE speakers showed the highest preference for indirectness. This study draws attention to the importance of the context in the analysis of democratization in non-Western societies, as the notion of face differs from the one dominating in the highly individualized Western world (also discussed by Hansen, 2018: 244). The paper concludes with a discussion on how linguistic directness does not need to be interpreted as a speaker's wish to emphasize their higher power position.

The same contrast between directness and indirectness in another outer-circle variety is found in Kotze and van Rooy's paper (this issue). Their study of modals and semi-modals in BrE, AusE and SAfrE also renders unexpected conclusions, which are directly linked to socio-political changes. The authors analyse a 7.2-million-word corpus including the British, Australian and South African Hansards, taken at five sampling points from 1901 to 2015, and study the verbs *must*, *should*, *HAVE to*, (*HAVE*) *got to*, *need* and *NEED to* along two semantic axes: (i) the expression of deontic, epistemic and dynamic meanings, and (ii) the strength of the obligation (median or strong). While the British and the Australian Hansards exhibit the declining frequencies of modal *must* and the increasing frequencies of less face-threatening verbs, the South African data clearly deviate from this pattern. Not only does *must* persist with a dominant deontic meaning, but also its frequency increases in the last period of the data; this increase is at least partially attributable to the entry of a new, confrontational political party (EFF) into the parliament, whose members use their voice to express their overt opposition in ways that would likely be seen as too direct in other varieties of English. The paper concludes with a reflection of the complex relation between political and linguistic democratization, where the former may actually foster the self-assertion of previously marginalized groups by means of face-threatening forms.

Finally, the role of directness in speech is also explored by Lundberg and Laitinen's (this issue) study of English as a Lingua Franca, as used by human Twitter users based in the Nordic countries and by Twitter trolls, i.e. bots located in Russia and Iran. This paper constitutes a novel approach to the study of democratization in at least two ways. Firstly, it applies advanced computational methods to the analysis of big data to provide two reliable and compact datasets, one containing 4.4 million words from the Nordic Tweet Stream (NTS) corpus, and the other containing 3.5 million words of troll data provided by Twitter. Secondly, it departs from the hypothesis that as the communicative purposes present in the troll data differ from those of human Twitter users, this is likely to be reflected in language use. The troll data is widely believed to represent anti-democratic forces aimed at shifting power balances in Western societies, whereas human-controlled accounts, although not completely democratic and harmonious, can nonetheless be assumed to aim at interpersonal communication. This hypothesis is confirmed by the results, which show clear quantitative differences between both datasets: troll messages are usually shorter, contain more negative particles and overuse features of formal registers, while genuine personal messages are more colloquial and spoken-like. The same pattern is found in a close analysis of the types of pronouns found in both datasets: while human-generated messages exhibit higher frequencies of first- and second-person pronouns (as in natural conversations), troll messages bear a closer resemblance to news regarding the high frequency of third-person pronouns. The paper constitutes an exploratory corpus analysis focusing on high-frequency elements such as pronouns, as they are expected to occur below the level of linguistic awareness and, therefore, constitute reliable evidence of the clear-cut differences between the language produced by humans and by (undemocratic) bots.

All in all, the seven corpus-based studies of the democratization of different varieties of English included in this issue contribute to the ongoing debate on linguistic/discursive democratization in different ways. To begin with, they contribute to disentangling the process from related ones, such as colloquialization and conversationalization. Secondly, they illustrate how different corpus techniques help unveil patterns that lead to common stylistic changes. Thirdly, they bring the concept of democratization to outer-circle varieties of English, spoken in societies where the concepts of equality and face may not

coincide with their Western interpretations. Finally, by putting the relevant changes in the socio-political context, they highlight the complex relation between language change and social and societal changes.

Acknowledgements

References

Alexander, Marc, Davies, Mark, 2016. The Hansard Corpus. https://www.english-corpora.org/hansard/.
Baker, Paul, 2010. Sociolinguistics and Corpus Linguistics. Edinburgh University Press, Edinburgh.
Barron, Anne, Schneider, Klaus P., 2008. Variational Pragmatics: A Focus on Regional Varieties in Pluricentric Languages. John Benjamins, Amsterdam.
Biber, Douglas, 1988. Variation across Speech and Writing. Cambridge University Press, Cambridge.
Biber, Douglas, 2010. Corpus-based and corpus-driven analyses of language variation and use. In: Heine, Bernd, Narrog, Heiko (Eds.), The Oxford Handbook of Linguistic Analysis. Oxford University Press, Oxford, pp. 159–191.
Biber, Douglas, Finegan, Edward, 1989. Drift and the evolution of English style: a history of three genres. Language 65 (3), 487–517.
Biber, Douglas, Gray, Bethany, 2012. The competing demands of popularization vs. economy: written language in the age of mass literacy. In: Traugott, Elizabeth Closs, Nevalainen, Terttu (Eds.), The Oxford Handbook of the History of English. Oxford University Press, Oxford, pp. 314–328.
Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, Finegan, Edward, 1999. Longman Grammar of Spoken and Written English. Longman, London.
Brezina, Vaclav, 2018. Statistical choices in corpus-based discourse analysis. In: Taylor, Charlotte, Marchi, Anna (Eds.), Corpus Approaches to Discourse. Routledge, New York, pp. 259–280.
Cameron, Deborah, 1990. Demythologizing sociolinguistics: why language does not reflect society. In: Joseph, John E., Taylor, Talbot J. (Eds.), Ideologies of Language. Routledge, London, pp. 79–96.
Clayman, Steven, John Heritage, 2002. The News Interview: Journalists and Public Figures on the Air. Cambridge University Press, Cambridge.
Conboy, Martin, 2010. The Language of Newspapers: Socio-Historical Perspectives. Continuum, London.
Conrad, Susan, 2004. Corpus linguistics, language variation, and language teaching. In: Sinclair, John (Ed.), How to Use Corpora in Language Teaching. John Benjamins Publishing Company, Amsterdam, pp. 67–85.
Culpeper, Jonathan, Nevala, Minna, 2012. Sociocultural processes and the history of English. In: Traugott, Elizabeth Closs, Nevalainen, Terttu (Eds.), The Oxford Handbook of the History of English. Oxford University Press, Oxford, pp. 365–391.
Diamond, Larry, 2015. Facing up to the democratic recession. J. Democr. 26 (1), 141–155.
Evert, Stefan, 2006. How random is a corpus? the library metaphor. Zeitschrift Für Anglistik Und Amerikanistik 52 (2), 177–190.
Fairclough, Norman, 1989. Language and Power. Longman, London.
Fairclough, Norman, 1992. Discourse and Social Change. Polity, Cambridge.
Fairclough, Norman, 2003. Analysing Discourse: Textual Analysis for Social Research. Routledge, London.
Fairclough, Norman, 2010. Critical Discourse Analysis: The Critical Study of Language. Pearson education, Harlow.
Farrelly, Michael, Seoane, Elena, 2012. Democratization. In: Traugott, Elizabeth Closs, Nevalainen, Terttu (Eds.), The Oxford Handbook of the History of English. Oxford University Press, Oxford, pp. 392–401.
Fruehwald, Josef, 2016. Filled pause choice as a sociolinguistic variable. Univ. Penn. Work. Pap. Ling. 22 (2), 6.
Gonçalves, Bruno, Loureiro-Porto, Lucía, Ramasco, José J., Sánchez, David, 2018. Mapping the americanization of English in space and time. PloS One 13 (5).
Grugel, Jean, 2001. Democratization: A Critical Introduction. Palgrave, Basingstoke.
Hansen, B., 2018. Corpus Linguistics and Sociolinguistics: A Study of Variation and Change in the Modal Systems of World Englishes. Language and Computers. Brill.
Hiltunen, Turo, Joe McVeigh, Säily, Tanja, 2017. How to turn linguistic data into evidence? In: Hiltunen, Turo, McVeigh, Joe, Säily, Tanja (Eds.), Big and Rich Data in English Corpus Linguistics: Methods and Explorations. Research Unit for Variation, Contacts, Helsinki. Studies in Variation, Contacts and Change in English.
Hiltunen, Turo, Räikkönen, Jenni, Tyrkkö, Jukka, 2020. Investigating colloquialization in the British parliamentary record in the late 19th and early 20th century. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.
Hundt, Marianne, Leech, Geoffrey, 2012. 'Small Is Beautiful': on the value of standard reference corpora for observing recent grammatical change. In: Traugott, Elizabeth Closs, Nevalainen, Terttu (Eds.), The Oxford Handbook of the History of English. Oxford University Press, Oxford, pp. 175–188.
Huntington, Samuel P., 1991. The Third Wave: Democratization in the Late Twentieth Century. University of Oklahoma Press, Norman, Oklahoma.
International Institute for Democracy and Electoral Assistance, 2019. The Global State of Democracy 2019: Addressing the Ills, Reviving the Promise. International IDEA, Stockholm. https://www.idea.int/publications/catalogue/global-state-of-democracy-2019?lang=en.
Janda, Laura A., 2013. Cognitive Linguistics: The Quantitative Turn: The Essential Reader. De Gruyter Mouton, Berlin.
Jockers, Lee, Matthew, 2013. Macroanalysis: Digital Methods and Literary History. University of Illinois Press, Urbana.
Jucker, Andreas H., Landert, Daniela, 2015. Historical pragmatics and early speech recordings: diachronic developments in turn-taking and narrative structure in radio talk shows. J. Pragmat. 79, 22–39.
Kopaczyk, J., Tyrkkö, J., 2018. Applications of Pattern-Driven Methods in Corpus Linguistics. John Benjamins Publishing Company.
Koplenig, Alexander, 2017. "The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets—reconstructing the composition of the German corpus in times of WWII. Digit. Scholarsh. Humanit. 32 (1), 169–188.
Kortmann, Bernd, 2018. Reflecting on the Quantitative Turn in Linguistics. Freiburg: University of Freiburg. https://bit.ly/2VlgFCL.
Kotze, Haidee, van Rooy, Bertus, 2020. Democratisation in the South African parliamentary hansard? A study of change in modal auxiliaries. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.
Kranich, Svenja, Hampel, Elisabeth, Hanna, Bruns, 2020. Changes in the modal domain in different varieties of English as potential effects of democratisation. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.
Laitinen, Mikko, 2007. Agreement Patterns in English: Diachronic Corpus Studies on Common-Number Pronouns. Société Néophilologique, Helsinki.
Landert, Daniela, 2014. Personalisation in Mass Media Communication: British Online News between Public and Private. John Benjamins Publishing Company, Amsterdam.
Leech, Geoffrey, 2004. Recent grammatical change in English: data, description, theory. In: Aijmer, Karin, Altenberg, Bengt (Eds.), Advances in Corpus Linguistics. Brill Rodopi, pp. 61–81.
Leech, Geoffrey, Fallon, Roger, 1992. "Computer corpora – what do they tell us about culture? ICAME J. 16, 29–50.
Leech, G., Hundt, M., Mair, C., Smith, N., 2009. Change in Contemporary English: A Grammatical Study. Cambridge University Press, Cambridge.
Lefkowitz, Julia, 2018. 'Tabloidization' or dual-convergence. Journal. Stud. 19 (3), 353–375.

Loureiro-Porto, Lucía, 2017. ICE vs Glowbe: big data and corpus compilation. World Englishes 36 (3), 336–365.

Lundberg, Jonas, Laitinen, Mikko, 2020. Twitter trolls: a linguistic profile of anti-democratic discourse. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.

Mair, Christian, 1997. Parallel corpora: a real time approach to the study of language change in progress. In: Ljung, Magnus (Ed.), Corpus-Based Studies in English. Rodopi, Amsterdam, pp. 195–209.

Mair, Christian, 2009. Twentieth-Century English: History, Variation and Standardization. Cambridge University Press, Cambridge.

Mair, Christian, 2019. American English: no written standard before the twentieth century? In: Yáñez-Bouza, Nuria, Moore, Emma, van Bergen, Linda, Hollman, Willem B. (Eds.), Categories, Constructions, and Change in English Syntax. Cambridge University Press, Cambridge, pp. 336–365.

Mollin, Sandra, 2007. The hansard hazard: gauging the accuracy of british parliamentary transcripts. Corpora 2 (2), 187–210.

Myhill, John, 1995. Change and continuity in the functions of the American English modals. Linguistics 33, 157–211.

Oakes, Michael P., Farrow, Malcolm, 2007. Use of the chi-squared test to examine vocabulary differences in English language corpora representing seven different countries. Lit. Ling. Comput. 22 (1), 85–99.

Palander-Collin, Minna, Nevala, Minna, 2020. Person reference and democratization in British English. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.

Paterson, Laura Louise, 2014. British Pronoun Use, Prescription, and Processing: Linguistic and Social Influences Affecting "They" and "He". Houndmills. Palgrave Macmillan, Basingstoke, Hampshire ; New York.

Rühlemann, Christoph, Martin, Hilpert, 2017. Colloquialization in journalistic writing: the case of inserts with a focus on well. J. Hist. Pragmat. 18 (1), 104–135.

Schneider, Edgar W., 2007. Postcolonial English: Varieties Around the World. Cambridge University Press, Cambridge.

Schützler, Ole, 2020. Frequency changes and stylistic levelling of though in diachronic and synchronic varieties of English – linguistic democratisation? In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.

Sinclair, John, 2005. "Corpus and text – basic principles. In: Wynne, Martin (Ed.), Developing Linguistic Corpora: A Guide to Good Practice. Oxbow books, Oxford, pp. 1–16.

Smith, Nicholas, 2003. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In: Facchinetti, Roberta, Krug, Manfred, Palmer, Frank (Eds.), Modality in Contemporary English. Mouton De Gruyter, Berlin, pp. 241–266.

Smith, Nick, 2020. Conversationalization and democratization in a radio chat show: a grammar-led investigation. In: Hiltunen, Turo, Loureiro-Porto, Lucía (Eds.), New Perspectives on Democratization: Evidence from English(es). Special issue in Language Sciences.

Staff of the Congressional Quarterly, 2013. In: Concise Encyclopedia of Democracy. Routledge, New York.

Szmrecsanyi, Benedikt, 2015. About text frequencies in historical linguistics: disentangling environmental and grammatical change. Corpus Linguist. Linguistic Theory 12 (1), 153–171.

Tognini-Bonelli, Elena, 2001. Corpus Linguistics at Work. John Benjamins Publishing Company, Amsterdam.