

<https://helda.helsinki.fi>

Person reference and democratization in British English

Palander-Collin, Minna

2020-05

Palander-Collin , M & Nevala , M 2020 , ' Person reference and democratization in British English ' , Language Sciences , vol. 79 , 101265 . <https://doi.org/10.1016/j.langsci.2019.101265>

<http://hdl.handle.net/10138/315758>

<https://doi.org/10.1016/j.langsci.2019.101265>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.



Person reference and democratization in British English

Minna Palander-Collin^{a,*}, Minna Nevala^b

^a University of Helsinki, Finland

^b Tampere University, Finland



ARTICLE INFO

Article history:

Received 19 June 2019

Received in revised form 21 November 2019

Accepted 13 December 2019

Available online 23 January 2020

Keywords:

Person reference

People words

Democratization

Big data

Small data

Keywords

Corpus

ABSTRACT

This article explores the interrelatedness of societal changes and changes in language practices. By using a combination of corpus linguistic and socio-pragmatic methods, we track diachronic changes in word patterns and interpret findings in the framework of democratization. The data comes from a small and representative corpus of British English (ARCHER-3.1) and from three “big data” sets (*Google Books*, *British Library Newspapers* and *The Economist*). We suggest that data triangulation, including sociohistorical contextualization, allows us to conclude that especially from the mid-nineteenth century onwards words signaling social status and referring to individuals have decreased and from the first decades of the twentieth century onwards words referring to collectivities of people have increased.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Exploring societal and linguistic change¹

Our work stems from the idea that social changes and changes in language practices work in tandem, and processes such as democratization can be observed in language. By using a combination of corpus linguistic and socio-pragmatic methods, it is possible to track social and cultural developments over long periods of time with an evidence-based approach (cf. [Farrelly and Seoane, 2012](#); see also [Hiltunen et al., this issue](#); [Smith, this issue](#)). In this study we focus on diachronic changes in word patterns and explore and interpret findings in the framework of democratization, which we understand broadly as “changing norms in personal relations” working towards less hierarchical and more equal patterns of social organization ([Leech et al., 2009: 259](#)).

Philologists, linguists and historians (e.g. [Hughes, 1988](#); [Williams, 1963, 1976](#); [Wierzbicka, 2006](#)) have established words and conceptual domains as important reflections of societal developments and cultural values. For example, [Hughes \(1988\)](#) identified moneyed words as culturally important and linked their development to the growth of capitalism. Linguists have focused on tracing more holistic patterns of twentieth-century language change as a reflection of broad societal trends such as colloquialization, Americanization, and democratization (e.g., [Leech et al., 2009](#); [Mair, 2006](#)). These studies use large corpora and quantitative corpus methodology to study grammatical changes. For example, shifts in the use of modal auxiliaries have been interpreted in the light of democratic developments and the levelling of power hierarchies ([Myhill, 1995](#)). Further

* Corresponding author. Department of Languages, University of Helsinki, PO Box 24 (Unioninkatu 40), FI-00014, Finland.

E-mail address: minna.palander-collin@helsinki.fi (M. Palander-Collin).

¹ This research was supported by a research grant no. 295383, Democratization, Mediatization and Language Practices in Britain, 1700–1950, Academy of Finland.

technological advancements in big data like *Google Books* and tools like Google Ngram Viewer have encouraged new types of efforts in mining huge amounts of lexical data to find out about human behaviour and cultural trends through the quantitative analysis of digitized texts (culturomics; e.g. [Pechenick et al., 2015](#)). We advocate for methodological triangulation, and our aim is to show that using different data sets and quantitative and qualitative methods including sociohistorical contextualization allows us to trace democratization processes in changing word patterns (for triangulation, see also [Kranich et al., this issue](#)).

We started our analysis with corpus linguistic tools using a relatively small but structured corpus (cf. [Humphreys et al., 2016](#)). Comparing consecutive centuries of British English data in [ARCHER-3.1](#) with keyword analysis, we observed several differences including changes in words referring to people. Although division by centuries is somewhat arbitrary, it seems that pre- and post-1900 data show differences that might be understandable in terms of broad democratization processes through which status differences are levelled out in linguistic expression. Some of the observed differences include the abundance of words referring to people in the pre-1900 data in comparison to the post-1900 data as well as qualitative differences in keywords. On the one hand, the eighteenth- and nineteenth-century people keywords refer to individuals and often signify social status (e.g. *lady, sir, captain, gentleman*). The twentieth-century keywords referring to people, on the other hand, include indefinite pronouns (e.g. *someone*) and words referring to groups of people (e.g. *workers, leaders, patients*).

Although [ARCHER-3.1](#) has been compiled to represent historical registers and to provide a solid basis for diachronic comparisons, it is possible that topics are not equally represented, which may affect keyword analysis. For example, [Oakes and Farrow \(2007\)](#) used a similar approach to compare the vocabularies of seven different ICAME corpora representing different regional varieties of English. Although they found clear differences between the corpora, they were cautious in attributing these differences in the words to cultural differences between the countries as the topics in the corpora may vary and there is no guarantee that the corpora represent typical topics discussed in that country (see also [Potts and Baker, 2012](#), for a similar comparison between American English and British English using the Brown family of corpora). Comparing similarities and differences between words in corpora with statistical methods entails many problems and insecurities, which have been discussed, for example, in [Kilgarriff \(2001\)](#) and more recently in [Lijffijt et al. \(2016\)](#).

To complement and verify the keyword analysis, we shall explore some of the keywords referring to people in their textual context in [ARCHER-3.1](#) and use big data including *Google Books* and *Gale Digital Newspaper Archives* to find out if the changes observed can be corroborated in other data sets and to see more specifically where the changes stem from. Changes in certain socio-pragmatic processes, as in social labelling and identification, often reflect changes in prevalent societal attitudes towards certain social groups and classes ([Bucholtz and Hall, 2005](#)). Similarly, these micro-level linguistic processes can trigger macro-level changes on their own through the process of democratization. Our research here is thus based on the crucial role socio-historical contextualization has in the understanding of these processes.

2. Democratization and language change

In political sciences democratization is typically understood in terms of specific societal changes through which “[l]arge numbers were given significant opportunity for influence over the political process” ([Garrard, 2002: 2](#)). According to [Garrard \(2002: 1–3\)](#), democratization in Britain was an ongoing process from the beginning of the nineteenth century and was “completed” around the mid-twentieth century. The end-products of the process include free, regular elections, in which all adults have the opportunity of influencing the political process, and liberal democracy, which entails e.g. freedoms of speech, association and press. Different social groups were admitted to the political process gradually so that the British electorate was prominently composed of the middling orders by 1832, the last middle-class men being admitted in 1918. Working-class men were included in the latter half of the nineteenth century through the Acts of 1867, 1884/5 and 1918, and the women’s formal parliamentary admission took place later in 1918 and 1928 ([Garrard, 2002: 3](#)).

The nineteenth century saw many intertwined legal, economic and societal changes that had an impact on how individuals were positioned especially in the labour market. More “liberal” legislation started to emerge in the 1830s and 1840s first protecting the most vulnerable groups like small children against harsh conditions in factories and mines. The mid-nineteenth century saw the first large-scale expressions of working-class discontent as well as continued trade union pressure. During the latter half of the century, the workers conditions improved. For example, the Trade Union Act in 1871 legalized trade unions in the UK for the first time, and the Employers and Workmen Act in 1875 put “masters” and “men” on equal footing regarding breaches of contract ([Atkinson, 2013; Frank, 2004; Hay, 2004](#)).

Industrial, commercial and urban growth go hand in hand in nineteenth-century Britain. Together with the industrialization of production, the increasing urbanization especially towards the end of the century severed the connection between home and work and increased anonymity and internal population movement ([Garrard, 2002: 273–274](#)). [Howkins \(1992 \[1991\]: 222–223\)](#) refers to the “‘modernisation’ of social relations” as a central change between 1850 and 1925. This entails the demise of the old order of “rank in which all had duties and obligations to those below and above of them” (223). Moreover, he maintains that “the poor and middling sort” were not passive in the process but rather active agents of social change, which shows for instance in the growing involvement in trade unions (223).

In linguistic studies the definition of democratization has been much vaguer and not directly linked to the study of political processes, although in broad terms we clearly talk about the same overarching phenomenon and its linguistic repercussions. [Fairclough \(1992: 98\)](#), for example, talks about democratization of discourse, which involves the reduction of overt markers of power asymmetries such as between teachers and pupils, or managers and workers. In the same vein, [Leech et al. \(2009: 259\)](#)

define democratization “as a reflection, through language, of changing norms in personal relations”. In much of linguistic research, democratization is understood as democratization of discourse, and Farrelly and Seoane (2012: 393) identify three (overlapping) areas of discursive democratization: “(1) the phasing out of overt markers of power asymmetry with the aim of expressing greater equality and solidarity (democratization proper); (2) a shift to a more speech-like style (colloquialization); and (3) a tendency toward informality in language, or informalization”.

In our study we see “the phasing out of overt markers of power asymmetry” in the keywords indexing individual social status and their decrease from the nineteenth to the twentieth century, and we think that these changes link to democratization processes in the political sense. Similar results as to the development of status words have been obtained e.g. by Leech et al. (2009: 259–260), who show that the use of titles *Mr*, *Mrs* and *Miss* preceding personal names in American English decreased from 1961 to 1992, or by Baker (2010) in the case of gendered words in twentieth-century British English from 1930 to 2005. Baker (2010: 145) concludes that the male bias still exists in 2005 as men are referred to more often than women, but the difference is diminishing. Another interesting finding in his study suggests that the inequality of the title system (*Mr*, *Mrs*, *Miss*, *Ms*) seems to be resolved in the decline of the entire system rather than *Ms* replacing *Mrs* and *Miss*.

In the light of our present results we could argue that such developments started even earlier in the nineteenth century. This is further supported by Vartiainen et al. (2019), who studied the development of the terms *master* and *servant* in three data sets: *The Old Bailey Corpus*, *The Hansard Corpus* and *the British Library Newspapers*. Their findings suggest that the use of these terms changed over the course of the nineteenth century, especially during the latter half of the century, reflecting legal and societal changes. However, the different data sources revealed somewhat different linguistic patterns and facets of social life, so that especially the *Hansard Corpus* reflected public discourse and meanings, such as the increasing use of *servant* in the sense ‘public servant’. The newspaper data, on the other hand, did not show this shift of meaning and overwhelmingly referred to ‘domestic servants’ as newspapers published small ads where (domestic) servants offered their services and were looked for by potential employers. This shows that the role of genre as a locus of practices should be kept in mind.

3. Person reference and people words

In its basic sense, reference is “a kind of verbal ‘pointing to’ or ‘picking out’ of a certain object or individual that one wishes to say something about” (Carlson, 2004: 76). In our study, we focus on the concept of person reference by way of looking at people words, which pertains here to the way in which speakers/writers refer to others (other-reference), such as the addressee or third parties who may be present to or absent from an interaction. When considering the traditional linguistic means through which expressions of reference can be accomplished, three main categories emerge: nominal forms (e.g. *Jim*, *Ms Miller*), pronouns (e.g. *we*, *they*), as well as other definite or indexical constructions (e.g. *this woman*, *the tall guy*). The group of nominal forms of reference comprises next to proper nouns also titles (e.g. *Professor*), kinship (e.g. *mum*, *dad*), as well as occupational terms (e.g. *doctor*, *reverend*). Person reference can also be established by the use of pronouns, both for individual (self-reference, agency) and collective reference (solidarity, group identity) (Fetzer and Bull, 2008).

Person reference as a merely deictic phenomenon has also been distinguished from membership categorisation (see e.g. Schegloff, 2007). Terms of membership categorisation mainly have a describing or identifying function by conveying what a specific person is ‘about’. These may include, for instance, terms pertaining to professional roles, such as *doctor* or *teacher*, which convey certain features relating to the activities people engage in as part of their occupation. While they may also be used for reference purposes, terms of membership categorisation are not necessarily equivalent to reference expressions depending on their context of use.

At the same time, the social, situational or professional role of an interlocutor may influence the referential expressions used. This means that the context and the participant’s role within that context have an effect on the choice of reference terms. They reflect socio-cultural conventions as well as personal characteristics and reveal insights into variables such as power, authority or distance as they mirror the interpersonal relationship between interlocutors (e.g. Nevala, 2009, 2019). Shifts in social and situational context can affect the appropriateness of referential expressions. Terms of reference therefore serve a variety of pragmatic functions in that they contain information about speaker and addressee, about their positions in society and, also, about their moods, intentions and relationship (Braun, 1998).

We know from earlier research that practices of person reference have undergone changes in the nineteenth century and these changes may be relevant in the light of our current study. Palander-Collin (2015) explored the genre of nineteenth-century newspaper advertisements and showed that person-mention was a prominent feature of the genre at the beginning of the period. Advertisements were often structured as if they were an interpersonal encounter between the advertiser and the audience showcasing deferential phrases familiar from correspondence and references to third parties as a way of establishing credibility. The ratio of advertisements with person-mention decreased towards the end of the period and many advertisements were then just product descriptions. Although it is difficult to say precisely what triggered this change, evidence from various primary data sets as well as historical research, i.e. data triangulation, pointing to a similar direction seems to speak for a shift in the reconceptualization of human relations in the course of the nineteenth-century.

So, person reference is intertwined with societal parameters and the interactants’ statuses within them, both at the interpersonal and societal levels of interaction. The group of person referential expressions, i.e. people words, that we

concentrate on in our current study comprises both individual reference terms denoting social status (e.g. *gentleman*, *lady*, *Sir*, *captain*) and collective terms denoting group membership (e.g. *workers*, *patients*, *leaders*).

4. Data and methods

In order to be able to focus on a diachronic continuum covering the period when democratization processes have identifiably taken place in Britain, the **ARCHER-3.1** corpus was chosen for keyword analysis. ARCHER is a multi-genre historical corpus of British and American English covering the period 1600–1999. The corpus has been designed as a tool for the analysis of language change and variation in a range of written and speech-based registers of English (for description of the corpus, see <http://www.helsinki.fi/varieng/CoRD/corpora/ARCHER/updated%20version/introduction.html>).

For the keyword analyses, we used only the British component from 1700 to 1999, which is about 1.07 million words of running text in eight genres including drama, fiction, sermons, journal or diaries, medicine, news, science, and letters. ARCHER is very small for studies based on lexical analysis, and for example the word *workers* that we explored was rare in the data. This is indeed one of the methodological drawbacks that ARCHER did not yield us more words referring to collective groups of people, and in further analysis it would be useful to identify more such words.

In order to see how specific statistical keywords were used in nineteenth century data outside of the ARCHER corpus and to counterbalance the possible problems of “small data”, the *British Library Newspapers* (Gale), *The Economist* (Gale) and *Google Books* (Google) were chosen as the data sources. These data sources provide access to a variety of British newspapers, both London-based and regional. All in all, the *British Library Newspapers* consists of the 17th and 18th Century Burney Collection Newspapers and the 19th Century Newspapers, which combined contain nearly 3 million pages and 40 million articles. *The Economist* has a continuous publication history from 1843 up to the present and it focuses on a selection of topics including international news, politics, business, finance, science, and technology. The databases by Gale are commercial, whereas the *Google Books* is freely available. These data sources incorporate a variety of genres and they represent public language use. As such they can be assumed to reflect current societal ideas and values.

These sources can be called “big data”, which entail many methodological insecurities as we do not know specifically, for example, which texts have been included and what the OCR quality of the texts is, or accessing the text behind the numbers produced by the search interface may be impossible. The use of the Gale databases is somewhat unwieldy, and as such they can be used primarily for qualitative work, but the tools are being developed and the search tools and statistics provided by the Gale interface were used in our study. As to the *Google Books*, the British component was used. *Google Books* have been used by other scholars, and, for instance, Laitinen and Säily (2018) provide a recent survey of this research. The pros and cons of small and big data have been discussed by Hundt and Leech (2012), who argue that small balanced and representative corpora are indispensable for corpus linguistics, and by Nevalainen (2013), who advocates data triangulation.

We identified statistically significant keywords in ARCHER comparing the nineteenth-century wordlist against the eighteenth- and twentieth-century lists and vice versa with WordSmith Keyword Tools. We used the log-likelihood test provided by the package (Scott, 2015: 245). The aim was to find out if keywords can reveal something about sociocultural shifts. In addition to comparing statistical trends in different data sets we wanted to explore what these statistical trends might mean in more detail in actual use. We singled out two central keywords, *gentleman* and *workers*, for closer inspection in the ARCHER data and mapped their semantic development and use in different genres.² These two words are not the most “key” according to the tests, but they seem to make sense in terms of sociocultural developments and changes in social relationships over time, briefly described above in Section 2. In further studies, other words could be explored; for instance, indefinite pronouns ending in *-one* seem to be characteristic of the twentieth century data, and their frequencies have been found to increase in other studies as well (Laitinen and Säily, 2018: 227 based on the *Google Books*). Finally, juxtaposing and comparing linguistic and historical findings is a central part of our methodological approach.

5. Results

5.1. Keywords referring to people in ARCHER

Table 1 summarizes all the keywords found when centuries in ARCHER were compared against each other. Positive keywords are those words that occur significantly more often in the data than in the comparison data, and the negative keywords stand for those words that occur significantly less frequently in the data than in the comparison data. The numbers in Table 1 are all the positive and negative keywords found using the log-likelihood test in WordSmith Tools, with settings looking for a maximum of 500 keywords occurring at least three times and in at least 5% of the files; the p-values are <0.000001.

² We did not look into the development of the plural form of the word *gentleman*, i.e. *gentlemen*, because it did not show up in the positive or negative keyword listings. As explained here, the listings showed a clear difference between the earlier use of terms in singular form denoting titles like *gentleman* and the later use of group membership terms in their plural form.

Table 1
Number of keywords in ARCHER (British).

	1800s vs 1700s	1800s vs 1900s
Positive keywords	66	81
Negative keywords	44	55
	1700s vs 1800s	1900s vs 1800s
Positive keywords	72	117
Negative keywords	37	60

Words referring to people include both nouns and pronouns and they were clearly present in the keyword listings as shown in Table 2. Keyword lists are of course not straightforward to interpret as some words have several meanings such as *major*. Others reflect institutions implemented in the nineteenth century (*police*) or relate to political upheavals in Europe (*Germans*). Proper names easily occur as *key*, but these have been excluded from calculations and further consideration as they reflect specific contents, such as characters referred to repeatedly in novels, rather than societal change. In this case, *sister* turned out to be a name-like word as it was used in a novel to refer to a nun on several occasions. Similarly, contracted forms are marked in Table 2, but they are not counted or considered further as they primarily reflect a change in spelling practices.

Table 2
Keywords referring to people in ARCHER (British) in order of keyness. Words in parentheses are excluded from further consideration.

	1800s vs. 1700s	1800s vs. 1900s
Positive keywords (% of all positive keywords)	(George), (I've), major, (I'm), police, (Robert), (you're), (Mary) (2/66; 3%)	Lady, my, sir, captain, thou, (Robert), his, I, me, (George), lord, prince, (am), cousin, (dear), mr, thy, countess, your (15/81; 19%)
Negative keywords (% of all negative keywords)	Person, him, king, I, gentleman , colonel, majesty, (John), (Peter), he, your, sir, my, her, his (13/44; 30%)	(William), sister, (he's), (Mary), (you're), (I've), (I'm), patients (2/55; 4%)
	1700s vs. 1800s	1900s vs. 1800s
Positive keywords (% of all positive keywords)	Madam, his, her, my, sir, your, mistress, he, (Peter), (John), esq, majesty, colonel, gentleman , I, king, him, person (16/72; 22%)	Patients, (I'm), (Arthur), (I've), (you're), (he'd), someone, Germans, workers , (Mary), (he's), leaders, sister, (Harold), anyone, (we've), everyone, (William), (we're) (9/117; 8%)
Negative keywords (% of all negative keywords)	(Mary), (Robert), (I'm), (Richard), (George) (0/37; 0%)	Your, thy, mr, (dear), (am), prince, lord, (Richard), (George), me, I, his, (Robert), captain, sir, my, lady (12/60; 20%)

We can say that *key* people words are different between the centuries. There are several people words in the eighteenth-century positive keywords list as opposed to the nineteenth century. In the eighteenth-century data vs. nineteenth-century data 22% of all the positive keywords refer to people, but in the reverse comparison only 3% of the positive keywords refer to people. The nineteenth-century positive keywords list vs. twentieth-century data again shows more keywords referring to people (19%) than the reverse comparison (8%). It seems then that talking about people is a characteristic feature of the eighteenth century but less so in the nineteenth and even less in the twentieth century.

We can also see other differences in Table 2 as eighteenth- and nineteenth-century data produce keywords that refer to individual persons (e.g. personal pronouns, *person*) and to a hierarchical social order (e.g. *king*, **gentleman**, *colonel*, *sir*, *madam*, etc.), whereas the twentieth-century data produce keywords referring to collective groups of people (*patients*, **workers**, *leaders*) as well as to unknown or generic persons with indefinite pronouns (*someone*, *anyone*, *everyone*).

5.2. Evidence from big data: Google Books, The Economist, and British Library Newspapers

One of the first questions that arises is whether we have detected a general diachronic trend in British English, or are we just uncovering characteristics of ARCHER data. In order to answer this question, three other available sources, *Google Books*, *The Economist* and *British Library Newspapers*, were used. The main trends are shown in Fig. 1–2³ and they are surprisingly and convincingly similar to the findings based on ARCHER.

Fig. 1a–c shows diachronic frequencies of words indexing individuals in the social order in the three big data sets. These words include *king*, *gentleman*, *colonel*, *majesty*, *sir*, *esq.*, *prince*, *lord*, *captain*, *lady*, *madam*, and *mistress*. Fig. 2a–c then show the words referring to groups of people in the same way: *workers*, *leaders*, *patients*. The words were picked from the keyword lists

³ We are grateful to Tuomas Salo for preparing the Figures.

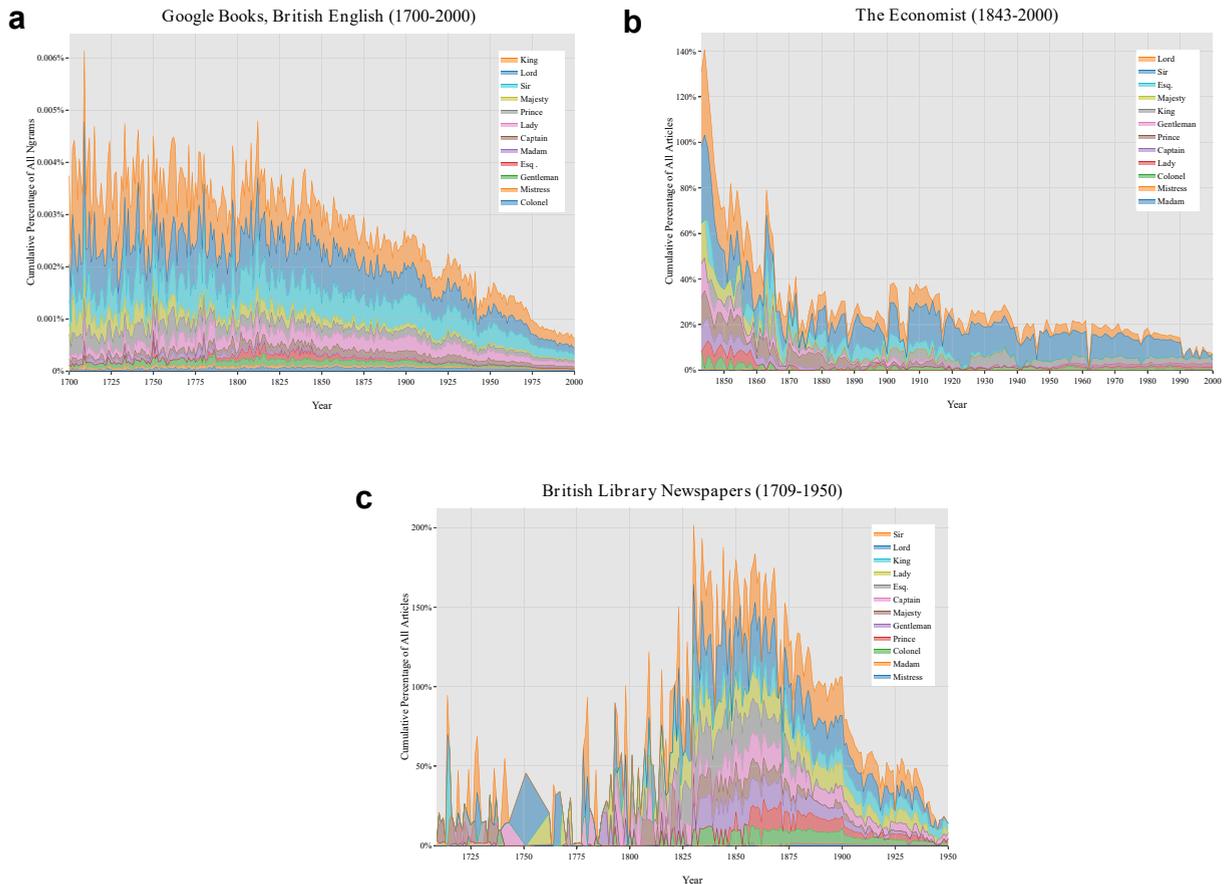


Fig. 1. a. King, gentleman, colonel, majesty, sir, esq., prince, lord, captain, lady, madam, and mistress in Google Books (1700–2000, British English). b. King, gentleman, colonel, majesty, sir, esq., prince, lord, captain, lady, madam, and mistress in *The Economist* (1843–2000). c. King, gentleman, colonel, majesty, sir, esq., prince, lord, captain, lady, madam, and mistress in *The British Library Newspapers* (1709–1950).

shown in Table 2. The timespans vary slightly and begin at 1700, unless the data set begins later or the first occurrences of keywords are later, and they end at 2000 or when the data set ends. In Fig. 1a–c the twelve words are in frequency order according to the highest peaks. In Fig. 2a–c the three words are in reverse alphabetical order. The scale of the Y-axis shows the cumulative percentage of the words of all articles for *British Library Newspapers* and *The Economist*, and cumulative percentage of all Ngrams for *Google Books*. This is because we have relied on the search tools and statistics provided by these services. Therefore, we are not suggesting a direct comparison of the data sets or of individual words, but we pay attention to the general trends. These big data sources show that references to collective groups increase (Fig. 2a–c) and status-related references to individuals (Fig. 1a–c) decrease as we saw in ARCHER. Moreover, the turning points in the timeline are similar.

References to individuals with status words decrease in all data sets most clearly after the mid-nineteenth century (Fig. 1a–c). The mid-nineteenth century has been identified as a watershed in many societal developments leading to more equal social relations (see Section 2), and these data sets seem to give further evidence to such an interpretation. However, our big data is also somewhat difficult to explain in this framework, as peaks in *British Library Newspapers* (Fig. 1c) seesaw in the eighteenth century without showing a clear pattern and indicate lower frequencies of these words than in the nineteenth-century data. The most likely explanation is missing or erroneous data, as from 1825 onwards the pattern in the *British Library Newspapers* data corresponds to graphs shown for *Google Books* (Fig. 1a) and *The Economist* (Fig. 1b). A closer inspection of the data sets might also reveal further genre differences and slight differences in the timing of the decrease. *The Economist* data (Fig. 1b), for example, indicates an earlier dip in the use of these words than the other two sources. The first major decline in *The Economist* seems to occur as early as 1850 and again in 1870, while the clearest dips in the *British Library Newspapers* occur in 1875 and 1900 and *Google Books* show a more even decline.

References to the groups of people start increasing a bit later in the early decades of the twentieth century, most clearly around 1910 in all data sets (Fig. 2a–c) and again more clearly after 1940 in the *Economist* (Fig. 2b). In isolation the three words can be assumed to relate to different topic areas - *workers* and *leaders* to political and industrial developments and *patients* to medicine and science - which may also be a factor behind each graph. *The Economist*, for example, has relatively few references to *patients* but plenty of references to *workers* and *leaders*, which makes sense in the light of the fact that one of the

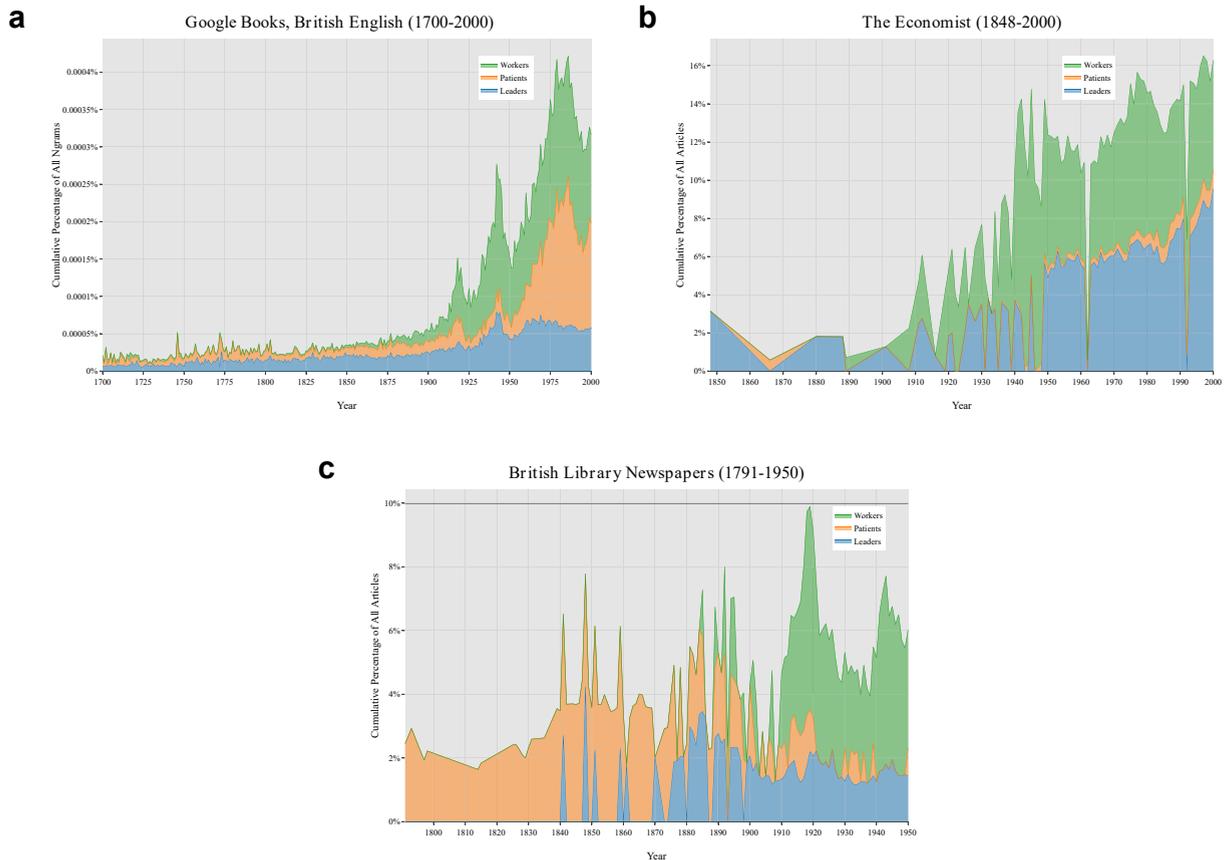


Fig. 2. a. *Workers*, *leaders*, and *patients* in Google Books (1700–2000, British English). b. *Workers*, *leaders*, and *patients* in *The Economist* (1848–2000). c. *Workers*, *leaders*, and *patients* in *The British Library Newspapers* (1791–1950).

major areas of focus of this publication is politics. In all three data sets *workers*, in particular, is an interesting case as the word starts occurring and gaining prominence after 1850, which links it to societal democratization processes. But, again the three sources show somewhat different dates: *workers* occur with any prominence in *Google Books* in 1850 (Fig. 2a), in *The Economist* in 1890 (Fig. 2b), and in the *British Library Newspapers* in 1880 (Fig. 2c).

5.3. Diachronic change and variation: analysis of gentleman and workers

In this section we review the development and use of two of the individual keywords to see how their use in context changed. As already mentioned, the eighteenth- and nineteenth-century people keywords and the twentieth-century keywords seem to differ according to reference type: earlier keywords refer to individuals and often signify social status (e.g. *lady*, *sir*, *captain*, *gentleman*), whereas those used later in the twentieth-century tend to include indefinite pronouns (e.g. *someone*) and words referring to groups of people (e.g. *workers*, *leaders*, *patients*). We have here opted to analyse the words *gentleman* and *workers*, as these most clearly link to changes in the significance of status hierarchies. Both of the terms also served as positive keywords in the data, *gentleman* for the nineteenth-century and *workers* for the twentieth-century material.

Table 3 shows the instances of the keyword *gentleman* in the data by genre categorization.⁴ We can see that it is mostly used in the fiction (167 instances in total; by normalized frequency the second frequent genre) and drama (115 instances; by normalized frequency the most frequent genre) categories in all time periods, and the use is primarily focused on the late eighteenth- and late nineteenth centuries. Other more prominent genres are news (46 instances; particularly in 1750–99), medicine (30 instances) and journals (26 instances). After the turn of the twentieth century, there is a sudden decrease in both the fiction and the drama, which confirms the trends also seen in the “big data” set in Fig. 1a–c. When we look at the development by 50-year time periods, the use of *gentleman* seems to be most prolific in the eighteenth century (216 instances

⁴ Table 3 shows data from the latter half of the seventeenth-century that was not used in the keyword comparisons in Section 5.1 as keyword comparisons were based on century vs. century comparisons.

in total; normalized frequency from 6.0 to 6.2), decreasing to only 25 instances during the twentieth century (normalized frequency from 1.0 to 0.4).

Table 3

The instances of *gentleman* and their normalised frequencies (per 10,000 words) according to genre in ARCHER.

	drama	fiction	journals	medicine	news	science	letters	total
1650–99	23 (8.6)	16 (3.9)	3 (1.4)	5 (2.2)	5 (2.2)	1 (0.5)	0	53 (2.9)
1700–49	15 (6.0)	61 (13.9)	0	8 (3.6)	11 (5.1)	5 (2.4)	6 (5.0)	106 (6.0)
1750–99	28 (11.7)	35 (7.8)	7 (3.2)	10 (4.8)	26 (11.3)	1 (0.5)	3 (2.5)	110 (6.2)
1800–49	17 (6.5)	11 (2.4)	9 (4.1)	6 (3.0)	1 (0.4)	5 (2.4)	0	49 (2.7)
1850–99	27 (10.2)	28 (6.5)	3 (1.3)	1 (0.5)	3 (1.3)	1 (0.5)	4 (3.7)	67 (3.7)
1900–49	3 (1.3)	12 (2.7)	2 (0.9)	0	0	0	0	17 (1.0)
1950–90	2 (8.2)	4 (0.9)	2 (0.9)	0	0	0	0	8 (0.4)
Total	115 (6.5)	167 (5.4)	26 (1.7)	30 (2.0)	46 (2.9)	13 (0.9)	13 (1.6)	410 (3.2)

As a term showing one's social status, *gentleman* was originally used of a man of the lowest rank of the English gentry, standing below an esquire and above a yeoman (Keen, 1990: 12, 2002). Gentlemen were considered members of the gentry proper until the sixteenth century, when people started to give new meanings to different titles, and the boundary between, for example, *gentleman* and *esquire* started to blur. In our data, there are clear instances of the use of *gentleman* as a marker of one's social status, particularly during the seventeenth and early eighteenth centuries. In examples (1)–(3), the term is used in various contexts, which all refer to status: in (1) a gentleman is juxtaposed with “Irish wretches”; in (2) mentioned alongside with “a vast estate”; and in (3) from Thomas Amory's novel, in reference to “the mountains of Richmondshire”, a part of the Yorkshire Dales historically famous for its manors and outlying properties.⁵

- (1) The manner of one families murder was thus; **A Gentleman** sent some Irish wretches to murder the family of an Englishman: (1653merc.n2b; *Mercurius Politicus*)
- (2) Among the rest of Sylvia's admirers was **one old gentleman** who was not much short of his grand climacteric, being turned of sixty, but of a vast estate [...] (1723blac.f3b; Arthur Blackmore, *Luck at Last; or the Happy Unfortunate*)
- (3) [...] this all madam, the thing that brought me here among the mountains of Richmondshire, was to find **a gentleman of my acquaintance** [...] (1756amor.f4b; Thomas Amory, *The Life of John Bunclie, Esq.*)

In the course of the seventeenth century, the notion of a gentleman started to change, when gentility began to be equalled with virtue, education and a capacity to govern (Palliser, 1983: 70). By the beginning of the eighteenth century, *gentleman* had developed a meaning different from the one denoting social status (Keen, 1990: 22, 2002). This meaning involved a certain superior standard of conduct, and sometimes appearance, and has since referred more to ‘behaving like a gentleman’ than to ‘being a gentleman’. Also in our data, we found that as the number of instances of *gentleman* decreases, the meaning of the term seems to be increasingly used for describing one's personal characteristics. Examples (4)–(5) show instances of this usage from the late nineteenth century, which describe the person's behaviour and general conduct.

- (4) There was a sense of Christmas about the travellers and the people who were at the terminus to meet them. The porter who came to the carriage door reminded Trefusis by his manner and voice that the season was one at which it becomes **a gentleman** to be festive and liberal. (1887shaw.f6b; George Bernard Shaw, *An Unsocial Socialist*)
- (5) [...] of his way to call the man “brother,” and to give him an opportunity of behaving like **a gentleman**, but his kindly forbearance had been wasted. (1896fred.f6a; Harold Frederic, *The Damnation of Theron Ware, or The Illumination*)

For the second people word, we chose to look more closely at *workers*, which also relates to societal developments depicted in Section 2. The overall number of instances in the corpus data was low, as can be seen in Table 4. The use of *workers* weighs heavily on the twentieth century, and on the news genre (26 instances in total; normalized frequency 1.6). Considering the reference data from the big data set, *The Economist* in particular, our findings seem to follow the general trend of *workers* being used more from the 1940s onwards. The corpus data are not of course sufficient for any further conclusions on more specific diachronic developments.

⁵ The first reference in parentheses gives the Archer file code, which includes the year of publication in the first four digits and the text type in the first letter after the full stop: f = fiction, n = news, j = journal, d = diary, m = medicine.

Table 4The instances of *workers* and their normalized frequencies (per10,000 words) according to genre in ARCHER.

	drama	journals	medicine	news	total
1650–99		1 (0.5)			1 (0.1)
1700–49					
1750–99					
1800–49					
1850–99					
1900–49				11 (5.6)	11 (1.4)
1950–90	1 (4.1)	1 (0.4)	2 (1.0)	15 (6.5)	19 (1.1)
Total	1 (0.1)	2 (0.1)	2 (0.1)	26 (1.6)	31 (0.2)

Despite the low occurrence rate of *workers* in the data, particularly when compared to that of *gentleman*, our searches confirm that the term is connected with group membership and occupational collectivities. There is a sense of anonymity and impersonalization, even objectification, in many cases. This is in sharp contrast with the notion and use of *gentleman*, which often relates to an active person and agent. In examples (6)–(10), the collective of *workers* is used in the context of, for example, war, battle, (human) rights and revolution, mostly connotating that the group of *workers* is something to protect or to protect against.

- (6) [...] the artillery to the opening of our trenches, to check some little insults the enemies sometimes attempted to make on **our workers**. (1691rich.j2b; Michael Richards, *Contemporary Diary of Siege of Limerick*)
- (7) Only Hope in **Allied Workers**. All believe firmly, however, that the will of **the British and French workers** will ensure an early revision of these and other unjustifiably harsh terms (1919dai1.n7b; *Daily Herald*)
- (8) Now they've broken into the shop downstairs. { = m WILLIAMS} No one can stop **the workers** from taking a few of their needs. { = f EMER.} Who's talking about their needs? (1958john.d8b; Denis Johnston, *The scythe and the sunset*)
- (9) Red Guard posters and leaflets in Peking, quoted by correspondents of Japanese and Czechoslovak newspapers, labelled **the workers of Nanking** as "the instrument of the bourgeois and reactionary Communist Party Committee of the province." (1967stm1.n8b; *The Sunday Times*)
- (10) In neighbouring Iran, there is concern for the safety of about **40 British mission workers** still in the country. (1979stm1.n8b; *Sunday Telegraph*)

The semantic diachronic differences closer to those in the use of *gentleman* mentioned above only show when we look at the singular of the term, *worker*. Examples (11)–(12) illustrate the variation between status/occupational standing and personal characteristic/behaviour. In (11), on the one hand, the term is collectively used both in the plural, "thousands of workers" and in the singular, "the local authority manual worker". A more personified meaning of the term can, on the other hand, be seen in (12), in which "a tremendous worker" is used to describe the person of Eddie Shackleton.

- (11) Some Ministers believe the Government must prepare itself to face a major public sector strike, possibly **the local authority manual worker**, if its credibility is to be restored. **Thousands of workers** in industry face the prospect of being laid off this week because of the rapidly worsening disruptive impact of the unofficial lorry and oil tanker drivers' strikes in many parts of Britain. (1979obs1.n8b; *Observer*)
- (12) I was glad to find that Eddie [Shackleton] too had shifted back to believing in an agreed solution and a Bill in the next Session. He showed me the huge new official paper he'd had prepared on the working of the two-tier system as well as his invaluable check list. He really is **a tremendous worker** and his Cabinet papers are some of the best I've ever seen. (1968cros.j8b; Richard Crossman, *The Diaries of a Cabinet Minister*, Vol. 2)

All in all, there is some indication, firstly, that the shift from terms used to express social status to collective group terms is a general diachronic trend. Secondly, the data show that there is semantic variation within terms like *gentleman* and *worker(s)*, depending on whether they are used to describe one's social standing or personal characteristic. Why this seems to be so and how the change from singular to collective references might relate to democratization are more closely discussed in the following section.

6. Democratization or register/genre development?

It is difficult to draw clear boundaries between motivations for changes observed in sections above. The decrease in words indexing individual social roles and often pointing to hierarchical relationships, such as *lady* or *gentleman*, can be understood as "democratization proper", i.e. the levelling of power asymmetries. It is less obvious, however, whether the observed increases in collective references to *workers*, *leaders* and *patients* are signs of democratization. In some ways, they could be, as at least *workers* and *patients* bring forth new types of social groups and they do not refer to the elites like the words indexing individuals we have discussed here. Democratization in the political sense also gave new groups of people, such as working-classes (*workers*), access to the political process and perhaps in that way more references start emerging.

Keywords and conceptual domains can be understood as important reflections of societal developments and cultural values. The people words that have occurred in our study as (statistical) keywords are not, however, necessarily recognized as cultural keywords per se by contemporaries and sociohistorians (e.g. Hughes, 1988; Williams, 1963, 1976). Instead, we might

talk about what Hughes (1988: 60–61) calls “the democratization of status-words”, which refers to the semantic shift as a whole, brought on by the break-down of feudalism. This process comprises of a change in understanding cultural status-words like *chivalry* or *courtesy* as being based on action instead of being derived from privilege of birth. Those words which were originally rooted in birth became more concerned with behavior, exactly what was discussed in relation to *gentleman* in Section 5.4. Another cultural change relates to the Industrial Revolution and the term *the masses*, which from the 1830s onwards was increasingly used to refer to people acting together, i.e. ‘common people’, ‘working people’ and ‘ordinary people’. In the twentieth century, keywords like *employers* replaced the earlier *masters*, only to be used variably with the term *the management* as an opposite to *workers* (Williams, 1976: 157–158, 161). This development of people words denoting individual action into those meaning group action can thus be clearly seen in our study as well.

Looking at the occurrences of people words more closely, we can identify other types of developments which might have affected the results. One of the most obvious ones involves genre. As can also be partly seen in Tables 3 and 4, *gentleman* and the other earlier people words denoting social status seem to occur mostly in fiction and drama texts, whereas the later collective ones such as *workers* appear to be clustered in the news category of the ARCHER. The spread of words like *workers* and *leaders*, for instance, might be connected to the development of the news genre in general and the increase in the nineteenth- and twentieth-century public debates of various ideological issues in particular. If we look at newspapers like *The Daily Herald*, and also later *The Daily Mirror* which was heavily commercialized, their early twentieth-century readership was built on and for the working classes (Conboy, 2010: 119–127). *The Herald* for one was founded to bring forth the workers’ perspective, being “the official exponent of the views of the great British Labour and Trade Union Movement” (Conboy, 2010: 120).⁶ It is no wonder then that words like *workers* and *leaders* were increasingly used in the news features in those newspapers which concentrated on these issues.

However, in order to try and focus on the effects of democratization, we also took a look at how people words in other registers than fiction, drama or news behave in the data. The genres that proved to be particularly interesting in this respect were the scientific and medical texts. As can be seen in examples (13) and (14), a change similar to that in the use of words like *gentleman* and *workers* seems to have happened here as well. In the first nineteenth-century example, the object of the medical examination is referred to as “a woman”. Example (14), typical of modern medical texts, collectively refers to a group of “2,381 patients” that have been the object of a wider survey. Naturally we have to take into account the difference between the two procedures discussed, but the data seem to indicate that the manner in which objects of medical examinations are talked about has changed from singular anonymity to plural anonymity.

- (13) It is not common to have an opportunity of examining the body of **a woman** dying from the anaemia produced by a fibrous uterine tumour. On this account, the dissection of the following case was regarded by me with much interest, but I did not expect that its importance would turn out to be so great as it did. (1864dunc.m6b; J. Matthews Duncan, The source of haemorrhage in a case of fibrous tumour of the uterus. *Edinburgh Medical Journal*, vol. 9)
- (14) In view of recent interest in the relationship of haematocrit and blood viscosity to hypertension and vascular disease, we have analysed retrospectively the relationship of haematocrit to blood pressure, vascular complications and other variables in 2,381 **patients** referred to the Glasgow Blood Pressure Clinic. (1985lowe.m8b; G.D.O. Lowe et al. Haematocrit in patients attending a hypertension clinic. *Scottish Medical Journal* 30:168–172)

Developments in various registers might have been boosters for the later increase in the use of collective people words. We are not certain as to what extent the corpus data support this explanation, however, and would like to emphasize the overall importance of the socio-economical need for new types of social groups, particularly from the 1850s onwards, which is more likely to have influenced variation within registers than the other way around.

7. Conclusion

In this study we set out to explore how the interrelatedness of societal changes, such as democratization, and changes in language practices could be observed in language. We used a combination of corpus linguistic and socio-pragmatic methods to see how changes in person reference in the form of people words relate to democratization processes, especially “democratization proper” (Farrelly and Seoane, 2012: 393). With keyword analysis of a small and representative corpus, we identified two interesting trends: words referring to an individual’s status decreased from 1700s to 1900s and some words referring to groups of people increased. In order to find out if this was an artefact of the data or a potentially significant trend, we used three “big data” sources to corroborate the findings. We also singled out two central keywords, *gentleman* and *workers*, to see how shifting trends were reflected in meanings and uses in context.

In the light of our analysis of word patterns as well as historical research findings, it seems that the mid-nineteenth century was a turning point when social relationships shifted in important ways. Of course, changes tend to be gradual rather than abrupt as different overlapping factors are at play. In the context of historical sociolinguistics and pragmatics it is important to pay attention to different genres as the locus of practices and changes that potentially take place at a different pace.

Methodologically, it seems that even smaller, well-structured, carefully-compiled corpora may be useful for pattern identification, but data triangulation proved particularly useful in our study. Solely relying on big data for pattern

⁶ *The Daily Herald* was relaunched as *The Sun* in 1964.

identification, on the other hand, might have left us wondering about the quality and specific contents of the data. As different data sets point to the same direction we can be fairly confident that practices of person reference have shifted.

References

Primary sources

ARCHER-3.1 = *A Representative Corpus of Historical English Registers* version 3.1. 1990–1993/2002/2007/2010/2013/2016. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Bamberg, Freiburg, Heidelberg, Helsinki, Lancaster, Leicester, Manchester, Michigan, Northern Arizona, Santiago de Compostela, Southern California, Trier, Uppsala, Zurich. Examples of Usage Taken from ARCHER Were Obtained under the Terms of the ARCHER User Agreement.
Gale, The Economist Historical Archive 1843–2015.
Google, 2013a. Google Books Ngram Viewer. <https://books.google.com/ngrams>.
Google, 2013b. About Ngram Viewer. <https://books.google.com/ngrams/info>.

Secondary sources

Atkinson, Evelyn, 2013. Out of the household: master-servant relations and employer liability law. *Yale J. Law Humanit.* 25 (2), 205–270. Available at: <https://digitalcommons.law.yale.edu/yjhl/vol25/iss2/2>.
Baker, Paul, 2010. Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gend. Lang.* 4 (1), 125–149.
Braun, Friederike, 1998. Terms of address. In: Verschueren, Jef, Östman, Jan-Ola (Eds.), *Handbook of Pragmatics*, Vol. 4. John Benjamins, Amsterdam. <https://doi.org/10.1075/hop.4.ter1>. Online.
Bucholtz, Mary, Hall, Kira, 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Stud.* 7 (4–5), 585–614.
Carlson, Gregory, 2004. Reference. In: Horn, Laurence, Ward, Gregory (Eds.), *Handbook of Pragmatics*. Blackwell, Malden, MA, pp. 74–96.
Conboy, Martin, 2010. *The Language of Newspapers: Socio-Historical Perspectives*. Continuum, London.
Fairclough, Norman, 1992. *Discourse and Social Change*. Polity Press, Cambridge.
Farrelly, Michael, Seoane, Elena, 2012. Democratization. In: Nevalainen, Terttu, Traugott, Elizabeth Closs (Eds.), pp. 392–401.
Fetzer, Anita, Bull, Peter, 2008. 'Well, I answer it by simply inviting you to look at the evidence.' The strategic use of pronouns in political interviews. *J. Lang. Politics* 7 (2), 271–289.
Frank, Christopher, 2004. Twelve: Britain, the defeat of the 1844 master and servants bill. In: Douglas, Hay, Paul, Craven (Eds.), pp. 402–421.
Garrard, John, 2002. Democratization in Britain: Elites, Civil Society, and Reform since 1800. Palgrave Macmillan, Basingstoke.
Hay, Douglas, 2004a. Two: England, 1562–1875. The law and its uses. In: Douglas, Hay, Paul, Craven (Eds.), pp. 59–116.
Hay, Douglas, Paul, Craven (Eds.), 2004b. *Masters, Servants, and Magistrates in Britain and the Empire, 1562–1955*. Chapel Hill, The University of North Carolina Press.
Hiltunen, Turo, Rääkkönen, Jenni, Tyrkkö, Jukka, 2020. Investigating colloquialization in the British parliamentary record in the late 19th and early 20th century. *Lang. Sci.* <https://doi.org/10.1016/j.langsci.2020.101270> this issue.
Howkins, Alun, 1992 [1991]. *Reshaping Rural England: A Social History, 1850–1925*. Routledge, London.
Hughes, Jeffrey, 1988. *Words in Time. A Social History of the English Vocabulary*. Basil Blackwell, Oxford.
Humphreys, Brendan, Nevala, Minna, Palander-Collin, Minna, Sairio, Anni, 2016. People, Work, Values: Tracing Societal Change Through Linguistic Shifts. A Paper Given in the Sociolinguistics Symposium 21, 15–18 June 2016. Universidad de Murcia.
Hundt, Marianne, Leech, Geoffrey, 2012. "Small is Beautiful". On the Value of Standard Reference Corpora for Observing Recent Grammatical Change. In: Nevalainen, Terttu, Traugott, Elizabeth Closs (Eds.), *The Oxford Handbook of the History of English*, pp. 175–188.
Keen, Maurice Hugh, 1990. *English Society in the Later Middle Ages. 1348–1500*. Harmondsworth, Penguin.
Keen, Maurice Hugh, 2002. *Origins of the English Gentleman: Heraldry, Chivalry and Gentility in Medieval England, c.1300–c.1500*. Tempus, Stroud.
Kilgarriff, Adam, 2001. Comparing corpora. *Int. J. Corpus Linguist.* 6 (1), 97–133. <https://doi.org/10.1075/ijcl.6.1.05kil>.
Kranich, Svenja, Hampel, Elisabeth, Bruns, Hanna, 2020. Changes in the modal domain in different varieties of English as potential effects of democratization. *Lang. Sci.* <https://doi.org/10.1016/j.langsci.2020.101271> this issue.
Laitinen, Mikko, Säily, Tanja, 2018. Zooming out. Overall frequencies and Google Books. In: Nevalainen, Terttu, Palander-Collin, Minna, Säily, Tanja (Eds.), *Patterns of Change in 18th-Century English. A Sociolinguistic Approach*. John Benjamins, Amsterdam/Philadelphia, pp. 221–241. 10.1075/ahs.8.c13.
Leech, Geoffrey, Hundt, Marianne, Mair, Christian, Smith, Nicholas, 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge University Press, Cambridge.
Lijffijt, Jeffrey, Nevalainen, Terttu, Säily, Tanja, Papapetrou, Panagiotis, Puolamäki, Kai, Mannila, Heikki, 2016. Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities* 31 (2), 374–397. 10.1093/dsh/ftq064, free access link: <https://academic.oup.com/dsh/article/31/2/374/2462752/Significance-testing-of-word-frequencies-in?guestAccessKey=b1a8cf86-2d64-4be6-b5c9-f60402df9799>.
Mair, Christian, 2006. *Twentieth-Century English. History, Variation and Standardization*. Cambridge University Press, Cambridge.
Myhill, John, 1995. Change and continuity in the functions of American English modals. *Linguistics* 33, 157–211.
Nevala, Minna, 2009. Altering distance and defining authority: person reference in Late Modern English. *J. Hist. Pragmat.* 10 (2), 238–259.
Nevala, Minna, 2019. *Two miserable creatures or those atrocious criminals? Evaluative reference in the Mannings murder reporting*. In: Lutzky, Ursula, Nevala, Minna (Eds.), *Reference and Identity in Public Discourses*. John Benjamins, Amsterdam, pp. 19–41.
Nevalainen, Terttu, 2013. English historical corpora in transition: from new tools to legacy corpora? In: Bennett, Paul, Durrell, Martin, Scheible, Silke, Whitt, Richard J. (Eds.), *New Methods in Historical Corpora*. Gunter Narr Verlag, Tübingen, pp. 37–53.
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. Linguistic Comput.* 22 (1), 85–99.
Palander-Collin, Minna, 2015. Changing genre conventions and sociocultural change: person-mention in 19th-century English advertisements. In: Bös, Birte, Kornexl, Lucia (Eds.), *Changing Genre Conventions in Historical English News Discourse*. John Benjamins, Amsterdam/Philadelphia, pp. 81–102.
Palliser, David M., 1983. *The Age of Elizabeth: England under the Later Tudors 1547–1603*. Longman, London.
Pechenick, Eitan Adam, Danforth, Christopher M., Dodds, Peter Sheridan, 2015. Characterizing the Google Books Corpus: strong limits to inferences of socio-cultural and linguistic evolution. *PLoS One* 10, 10. e0137041.
Potts, Amanda, Baker, Paul, 2012. Does semantic tagging identify cultural change in British and American English? *Int. J. Corpus Linguist.* 17 (3), 295–324.
Schegloff, Emanuel A., 2007. Categories in action: person-reference and membership categorization. *Discourse Stud.* 9 (4), 433–461.
Scott, Mike, 2015. *WordSmith Tools Manual Version 6.0*. Stroud, Gloucestershire, UK. Lexical Analysis Software Ltd [Available online at [lexically.net](http://lexically.net/downloads/version6/wordsmith6)]
Smith, Nichol, 2020. Conversationalization and democratization in a radio chat show: a grammar-led investigation. *Lang. Sci.* <https://doi.org/10.1016/j.langsci.2020.101269> this issue.

- Vartiainen, Turo, Palander-Collin, Minna, Hiltunen, Turo, 2019. Master and Servant: Exploring Democratization in British Newspaper Texts, 1800–1925. A Paper Presented at the 7th International Conference on Historical News Discourse (CHINED VII), 23–24 May, 2019. Pompeu Fabra University.
- Wierzbicka, Anna, 2006. *English: Meaning and Culture*. Oxford University Press, Oxford.
- Williams, Raymond, 1963 [1958]. *Culture and Society 1780–1950*. Penguin Books, Harmondsworth.
- Williams, Raymond, 1976. *Keywords: A Vocabulary of Culture and Society*. Fontana, London.