

Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English

Paul Baker

Abstract

In order to investigate frequency and context of usage of gender marked language, four equal sized and equivalently sampled corpora of British English in a range of written genres (press, fiction, general prose, learned writing), from 1931, 1961, 1991 and 2006 were compared. Terms that were investigated included male and female pronouns, man, woman, boy and girl, gender-related profession and role nouns such as chairman, spokesperson and policewoman, and terms of address such as Mr and Ms. Some reductions in frequencies of male terms were found over time, particularly in terms of decreases of male pronouns and Mr. However, equal frequencies did not necessarily equate with equal representation. A qualitative analysis of man and woman found that while there had been some reductions in gender stereotypes, others were being maintained (such as a lack of adjectives like successful or powerful being applied to words like woman). Additionally, the term girl was still more likely than the term boy to refer to adults, and it was often used in a disparaging or sexual way. The article concludes with a discussion of the sort of linguistic strategies that appear to have been successful in terms of equalising gender representation.

KEYWORDS: CORPUS; GENDER; REPRESENTATION; DIACHRONIC; VARIATION

Affiliation

Department of Linguistics and English Language, Lancaster University, UK.
email: p.baker@lancaster.ac.uk

Introduction

Corpus linguistics involves the analysis of large collections of computerised texts, often carefully sampled in order to be representative of a particular language variety (McEnery and Wilson 1996, Kennedy 1998, Hunston 2002). Such texts (known as corpora) are usually subjected to quantitative forms of analysis via software which can identify frequent linguistic patterns. Corpus software can also identify frequent non-fixed combinations of words (collocations) and distinctive words, called 'keywords' by conducting comparisons between multiple corpora. Such patterns can be automatically identified via a naïve or *corpus-driven* approach – the analyst does not know what will be found in the corpus and starts with no hypotheses, but instead frequent or salient patterns become the focus of the analysis once they are uncovered. On the other hand, with a *corpus-based* approach, the analyst wants to use the corpus to explore a pre-determined hypothesis, topic or linguistic category. The corpus-based/driven distinction can be attributed to Tognini-Bonelli (2001), although later researchers (e.g. McEnery et al. 2006: 8), have argued that, as it is difficult to approach a corpus from a completely naïve stance, such positions can perhaps be thought of as extremes on a continuum. In fact, much corpus research tends to combine both approaches to various degrees.

As argued in Baker (2005: 5–14), linguists who are interested in language and gender have tended to eschew corpus approaches in the past, often preferring to use more qualitative, small-scale analyses. This could be due to a number of factors: unfamiliarity with the software and processes required to build and analyse corpora, a shift in social sciences epistemology to focus more on qualitative analysis coupled with dislike of a paradigm which puts emphasis on the notion of comparative 'difference', or a misconception of corpus linguistics as 'number-crunching' which does not take into account context. Some of these concerns are valid, although most can be overcome. As Biber (1988: 4) points out, corpus linguistics relies on both qualitative and quantitative methods: 'Association patterns represent quantitative relations, measuring the extent to which features and variants are associated with contextual factors. However functional (qualitative) interpretation is also an essential step in any corpus-based analysis.'

In the past, a small number of studies have used corpus methods, combined with a variationist approach to sociolinguistics in order to investigate differences between male and female speech. For example, Rayson et al. (1997) used 4.2 million words of transcribed speech from the spoken demographic section (consisting of private conversations) of the British National Corpus (BNC), using chi-squared tests to identify which words were distinctive of male and female speech. Typical male words included numbers, forms of the taboo word

fuck, agreement/disagreement and discourse makers: *no, yeah, aye, right*, articles and determiners: *the, a, that, which*, and colloquial terms for males: *mate, guy*. Typical female words included pronouns: *she, her, I, him, he, me*, certain adjectives (see Lakoff 1975): *nice, lovely*, discourse markers: *oh, mm* and words to do with speech and thought: *said, thought, think*. Kilgarriff (2001) used the Mann Whitney test (which compares rankings rather than actual frequencies) on the same data, and found stronger differences in lower frequency words such as *record, shot* and *square* for males, and *children, clothes, dish* and *shopping* for females. Schmid (2003) took a corpus-based approach to the spoken section of the BNC, looking for differences in particular domains which other authors have claimed to be indicative of gender differences. He found that females used more words relating to colour, the home and clothing, while males used more words to do with public affairs and abstract concepts. However, some of the gender stereotypes he investigated indicated a more complex picture: women used certain swear words more than men (including *shit* and *damn*), while men utilised certain hedges more often than women (*maybe, perhaps* and *sort of*).

While it is seductive to comment on differences, many corpus studies actually find similarities too, so such research should not be used in order to reify a 'gender differences' model of language use. Rayson et al. (1997) note that the differences they found only reflect tendencies not absolutes, while Schmid (2003: 219) concludes that for men and women, 'to a very large extent, these two 'cultures' overlap' (my quotes). However, while we should not over-generalise actual differences, I agree with Sunderland (2004: 16–7) who writes 'I do not (have not been able to) abandon the idea of gender as premised on 'difference', nor do I wish to, since it is important not to lose sight of the ways in which *notions of gender* can adversely affect women's access to important linguistic resources and possibilities of expression...' (my emphasis). Sunderland argues that society tends to (over-)focus on gender difference, which goes on to impact on people's lives.

A different strand of corpus studies on gender, which reflect Sunderland's point are those which have focussed on representation of gender, instead of comparing male/female usage. Rather than counting the frequencies of words in male and female speech (or writing), these studies have looked at the ways in which males and females are written or spoken about in corpora. For example, Kjellmer (1986) found that terms for males were more frequent than female terms in British and American written corpora from the early 1960s. Biber et al. (1999) examined the Longman English Corpus, reporting that 620 nouns end in the suffix *-man*, whereas only 38 end in *-woman*. Romaine (2001) examined the entire 100 million-word British National Corpus, finding that *Mr* occurs more often than *Mrs, Miss* and *Ms* combined and that terms that potentially erase females such as *chairman* and *spokesman* tend to be prevailing.

A detailed study by Pearce (2008) examined the BNC from a somewhat different perspective. Pearce looked at collocates (words which tend to co-occur next to or near another word with statistically significant regularity) of the lemmas MAN and WOMAN.¹ He used the corpus analysis tool Sketch Engine (Kilgarrieff and Tugwell 2002) in order to distinguish which verbs collocated with MAN and WOMAN when they occurred as subject or object. He found that women tended to take the object of verbs which denoted sexual violence, coercion and observation such as *rape*, *categorise*, *exhibit*, *monitor*, *regulate* and *define*. Women were the subject of verbs which constructed them as irritating: *fuss*, *nag*, *cluck*, *taunt*, *annoy* and *berate*. On the other hand, men were both the subject and object of non-sexual violence verbs, collocating with words like *oppress*, *pounce*, *raid*, *ransack* and *betray*. Men were also the subject of seduction verbs like *bewitch*, *captive*, *charm* and *flatter*. Taken together, most of the corpus studies of gender representation paint a depressingly sexist and stereotyping picture of the English language in the twentieth century. A slightly more encouraging study is Sigley and Holmes (2002) who found that there had been a move towards non-sexist language when they compared corpora from the early 1960s and the early 1990s. They reported reductions in the use of the pseudo-polite *lady*, the generic *man*, and the *-ess* and *-ette* suffixes. Additionally, Baker (2008: 38) compared written texts from three time periods in the BNC (1960–1974, 1975–1984 and 1985–1993), finding that there had been some reductions in sexist terms like *mankind* and *ladylike* and increases in use of non-sexist terms like *him or her*, *he or she*, *Ms*, *chairperson* and *police officer*. However, terms like *early man* and *male nurse/prostitute/stripper/model* had actually increased over the time periods examined. While of interest, these two studies have taken their most recent data from the early 1990s and can tell us little about contemporary language use. The purpose of this study was to determine whether trends regarding gender representation that appear to have been underway in the early 1990s are continuing into the present day. A potential criticism of studies that have only examined two (or even three) points in time, is that changes may not be especially reflective of ongoing change, but may in fact be due to chance. Sigley and Holmes' (2002) study used two sets of corpora which were separated by a 30 year period. Such corpora may provide 'snapshots' of language use in 1961 and 1991, but they do not directly tell us anything about the years in between. Theoretically, a word like *lady* could have risen and fallen in frequency at several points between 1961 and 1991. Therefore, sampling two points with a 30 year gap between them is not an ideal way of investigating diachronic language change. The present study aims to compare four corpora (two which have only recently been created), over a wider time span in order to build on and update existing corpus-based research on gender representation.

Method

For the purposes of this study, four corpora were utilised. All corpora were of equal sizes (a million words) and followed the same sampling framework, containing 500 samples of texts (each approximately 2,000 words in size) from four main categories (press, general prose, learned writing and fiction) of published writing in British English, further subdivided into 15 subcategories (romantic fiction, science fiction etc). All four corpora therefore follow the well-known Brown Corpus model. The Brown Corpus (initially called The Standard Corpus of Present-Day American English) was the first reference corpus ever created (Francis and Kučera 1979). Between 1970 and 1976 a British version of the Brown corpus was built, called the LOB (Lancaster-Oslo/Bergen) Corpus. This corpus contained texts from the same time period as Brown (1961). Two further corpora were later created at Freiberg University, using data from the early 1990s: Freiberg-Brown and Freiberg-LOB (often abbreviated to Frown and FLOB). Collectively, these four corpora have been referred to as the Brown Family. In order to remove culture as a complicating factor, only the British components (LOB and FLOB) are used in this study.

The British branch of the Brown family has had a number of recent additions. At Lancaster University, a 1930s version (called Lancaster 1931 or BLOB (Before-LOB)) has been built while a 1901 version is also underway. I have built a contemporary version of LOB, using texts that had been first published in paper form and then placed on websites (Baker 2009). Part of the motivation for building this contemporary corpus (called British English 2006 or BE06) was to investigate the feasibility of using the internet as a way of collecting corpus data. With the other corpus building projects in the Brown family, text had to be typed or scanned in by hand. The BE06 corpus was much easier to create as all of the texts already existed in electronic form (it took about twelve working days to build). It should be noted that the difference between BLOB and LOB, and LOB and FLOB is 30 years. However, the difference between FLOB and BE06 is only 15 years. As argued earlier, closer sampling points will help to ensure that a better image of trends over time can be extrapolated. Ideally, the Brown family should be added to every five years.

The corpora were uploaded to a web-based interface called CQPweb (Corpus Query Processor) which was developed by Andrew Hardie and Sebastian Hoffmann. The interface is freely accessible via a user-name and password.² Users can carry out concordances, perform frequency counts and examine distributions of linguistic items across genres or individual texts but do not have full access to each text in full. This is one way of circumnavigating issues of permissions and copyright which are likely to make large corpus-building

projects increasingly expensive and time-consuming. Mark Davies has also used this approach to make a number of large reference corpora freely available (including the Corpus of Contemporary American English, the BNC and the TIME magazine corpus).³

The following two research questions were born in mind while undertaking analysis:

What do frequencies in male and female pronouns and nouns across the four diachronic corpora suggest about male bias in language use?

How are males and females qualitatively represented across the four corpora?

Using CQPweb it was possible to obtain frequencies of specific nouns and pronouns in the four corpora. Then concordance analyses could be carried out in order to obtain a more qualitative-based picture of how such words were being used in context. One finding which arose from the analysis is that two equivalent words may have equal frequencies, but be used in different contexts. Frequency is therefore only one possible indicator of bias.

Analysis

The analysis is corpus-based rather than corpus-driven, in that the words chosen for examination were decided on by the analyst, rather than emerging from say, a top 100 list of frequent words. Such an analysis could be criticised in that certain words may be overlooked. Researcher intuition, coupled with examination of other research on gender-marked language use was helpful in determining a list of words that could be investigated. However, the caveat that the analysis is not an exhaustive account of all gender-marked language use is necessary.⁴

The analysis begins by focussing on gender-marked pronouns and then moves on to look at various nouns which refer to males and females. Because the corpora under examination are relatively small (a million words each), it is likely that we can make stronger conclusions about words which are frequent than those that do not occur very often. Some words which were initially chosen for examination (*ladylike*, *fireman*) were too infrequent to be worthy of comment. This suggests one potential drawback of using small corpora – they are better at telling us about trends in a few high frequency words rather than revealing a great deal about many low frequency words. With that said, a small number of high frequency words account for a great deal of our language use. Taken together, in the four corpora, just 380 words accounted for 62% of the total language. Additionally, the fact that all four corpora contain equal

amounts of language and followed the same sampling frame, means that comparisons can be directly made and concerns about multiple factors influencing the results are much reduced.

Another potential limitation of this study is that the Brown family reflect written and published rather than spoken or unpublished English. The texts are likely to have been written by adult, professional, middle-class authors (who tend to have more opportunities to have their writing published than other types of people). Any conclusions we make about language change and gender over time, therefore must be made with the proviso that they are not representative of all language users in the UK (although the high profile of these texts may influence other language users). Finally, it should also be considered that all of the texts from the BE06 were collected from the world wide web. While they were originally published in paper form, the fact that they were archived on the internet may have had an impact on the way that they were written. An initial investigation of the BE06 found that the rankings of its most frequent vocabulary items differ slightly from the other three corpora. So while the BLOB, LOB and FLOB share a Spearman similarity coefficient⁵ of 0.99 or 0.98 in terms of the rankings of their top 20 words, this figure is slightly lower at between 0.93 and 0.96 for the BE06 corpus when compared against the other three corpora. This measure indicates that the BE06 is most similar to the 1991 FLOB Corpus and least similar to the 1931 BLOB Corpus, however, which is what would be expected.⁶

Figures 1 and 2 show the frequencies at the four sampling points for male and female pronoun usage, while Figure 3 gives a direct comparison between all male and female pronouns. While the figures show lines in order for trends to be more easily seen, it should be noted that the lines themselves are projections of trends – there is only actual data for four sampling points (1931, 1961, 1991 and 2006). It can be seen that there have been decreases in usage of all male pronouns (at least since 1961), although this is most marked for the most frequent pronoun *he*. Female pronouns seem to show a slight increase, although this seems to be tailing off in the more recent data. The gap between male and female pronouns still exists and is substantial, but it appears to be shrinking. The data for female pronouns does not show a single clear pattern however. While *her/hers*⁷ shows a slight increase, *she* has actually appeared to decrease between the last two sampling points. When looking at other types of pronouns such as *I*, *you* and *they* (not shown as figures), it was found that in general there had been increases in first and second person pronouns over time, which is perhaps reflective of written English becoming more conversational and personalised in recent years (see Fairclough 1989, 1994). This may therefore relate to the slight decline in *she* and the male pronouns.

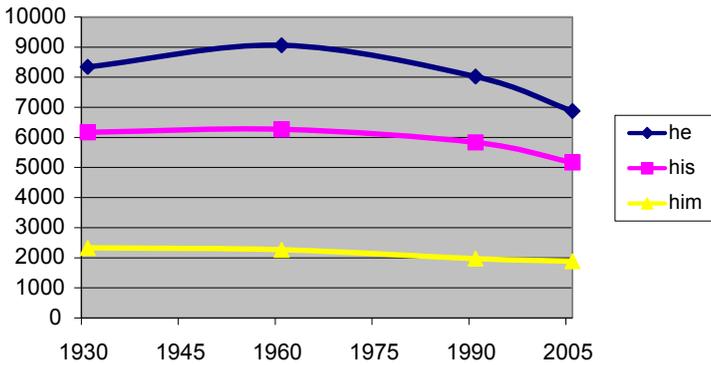


Figure 1: Frequency of male pronouns

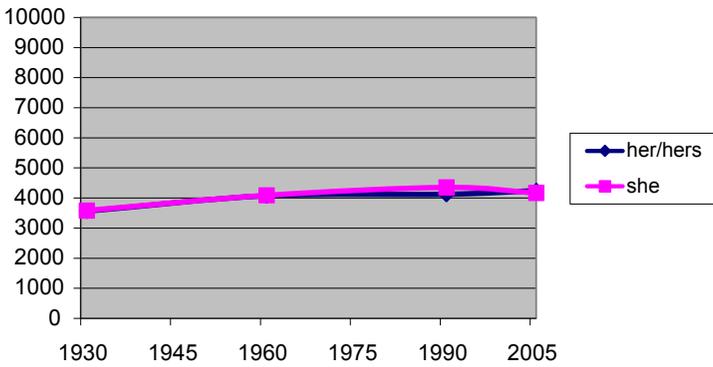


Figure 2: Frequency of female pronouns

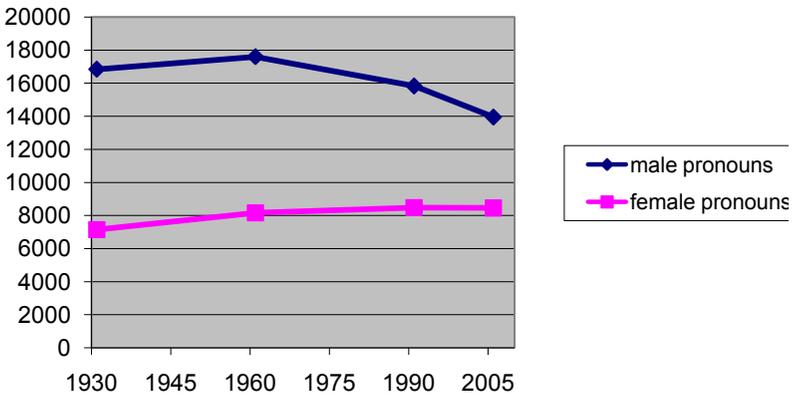


Figure 3: Frequency of all pronouns

Table 1: Frequencies of gender inclusive pronouns

	BLOB	LOB	FLOB	BE06
him or her	0	4	5	3
he or she	4	5	14	7
he/she	0	0	2	3
s/he	0	0	8	1
him/her	0	0	3	0
totals	4	9	32	14

Another aspect of pronouns involves the use of terms which attempt to be inclusive by presenting both alternatives, such as *him/her* or *he or she*.⁸ Table 1 gives the frequencies of these terms and similar ones. Apart from *s/he* which may be argued to place (part of) the female pronoun first, in general, there were no terms in any of the corpora which put the female pronoun first such as *she or he*. Therefore even attempts at inclusiveness could be criticised as referencing ‘male firstness’ (see Sunderland 1996: 94). As Hartman and Judd (1975: 390) argue, this ‘reinforces the second-place status of women and could, with only a little effort, be avoided by mixing the order’. Looking at Table 1, it is clear that while these uses of pronouns are rare, there seems to have been an increase between 1961 and 1991. However, the total for 2006 is less than half that of 1991, suggesting that this strategy is not becoming popular, and may even die out. Schwartz’s (2006) study of focus groups suggests that some people have trouble with some gender neutral terms like *s/he* because they find them distracting or messy. Such terms are likely to be easier to implement in writing rather than speech, which could prove to be a barrier to their long-term uptake.

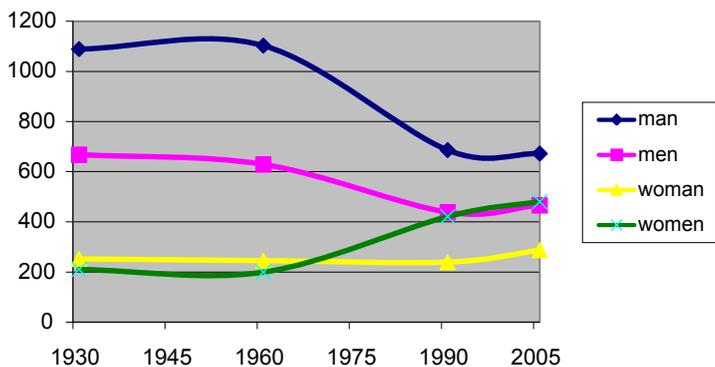


Figure 4: Frequencies of MAN and WOMAN

Moving on to consider nouns, Figure 4 shows the frequencies of *man*, *men*, *woman* and *women*. Overall these terms are used less frequently than gendered pronouns. However, what is interesting about the pattern of this figure is that there appears to be signs of convergence of all four terms. This is especially the case for *men* and *women*, which have almost the same frequencies in the BE06 data. Figure 4 does not consider cases of affixation though. Do the conclusions made by Romaine (2001) regarding the preference for words like *spokesman* and *chairman* still hold? Table 2 shows the frequencies of words that begin with *spokes-* (the table includes plural forms). First, it is noteworthy that the gender-neutral *spokesperson* is the least popular form. Second, there does appear to have been some uptake of *spokeswoman* and *spokesperson* since 1991, although the most widely used term is still *spokesman*. One question to consider, is whether *spokesman* is actually being used to refer to females. A concordance analysis (which examined all cases of *spokesman* in all four corpora in the contexts that they appeared in) found only one case of this in the BE06 data (out of 43 cases), so while this is a possible strategy, it is very rare.

Table 2: Frequencies of words starting with *spokes-*

	BLOB	LOB	FLOB	BE06
spokesman	1	22	50	43
spokeswoman	0	0	8	5
spokesperson	0	0	2	4

A further question could be raised regarding which strategy would actually be preferable in terms of gender equality? Would it be desirable to use *spokesperson* to refer to everyone, as this is a term which does not mark gender, thus signifying that everyone is capable of fulfilling the (important) role it refers to? Or would it make sense to explicitly mark women as carrying out the role, when they do so, by using *spokeswoman*? If we only use one term, is it acceptable to use *spokesman* to refer to a woman, or is this a form of erasure? It could be asked (somewhat radically) why not use *spokeswoman* instead to refer to everyone, male and female? Should multiple strategies be used or should some attempt at consensus be reached? An issue which arises here is that the frequencies of these words probably reveal a social reality – *spokesman* could be more frequent in all four corpora because it is likely to be a role which men tend to carry out more often than women.

Table 3: Frequencies of words referring to the police

	BLOB	LOB	FLOB	BE06
policeman	30	15	39	30
policewoman	0	3	0	3
WPC (woman police constable)	0	0	3	0
PC (police constable)	0	1	7	2
police constable	0	2	2	2
police officer	2	5	11	34
cop/copper	0	6	10	37

The frequencies of words which refer to the police are shown in Table 3. Both plural and singular forms were examined. Additionally, some terms occurred as one word, while in other cases they were written as two words, so the figures in the table are for both *police man* and *policeman*. Compared to *spokes-* there are a wider range of possible terms that can be used. In 1931 and 1961, the term *policeman* is preferred, although there is evidence that other terms are starting to be taken up. This pattern seems to continue in the 1991 data with *policeman* still most popular, but citations of *WPC*, *PC*, *police constable* and *police officer* occurring. Finally, in 2006, there appears to be a breakthrough with *police officer* becoming more popular than *policeman*. However, by this point, interestingly, the informal (and ostensibly gender neutral) *cop/copper* is the most frequent term to refer to the police. This term is sometimes used to refer to females e.g. ‘Rumours are, we even had one cop show her support on our petition!’ (BE06 F27). The 2006 corpus is the only one where gender-neutral terms for the police are more frequent than gender-marked terms. Again, similar questions could be raised. Is *police officer* a less sexist strategy than one which explicitly marks gender? It could be argued that the term *WPC* is somewhat sexist as there is no equivalent male term (*MPC?*). So the default term (*PC*) would be understood to refer to males, while the marked, exceptional case (*WPC*) is female. Again, the frequencies of terms, especially the increases in terms referring to women or that are gender-neutral could reflect a social reality. In the UK, there have been efforts to ensure that more women are employed in the police force, by using targets and positive discrimination (see Brown et al. 2006). Additionally, The British Association of Women in Policing was formed in 1987, describing their mission as ensuring that women of all ranks and grades in the police are heard.

Table 4: Frequencies of words starting with *chair-*

	BLOB	LOB	FLOB	BE06
chairperson	0	0	1	0
chairman	116	117	109	75
chairwoman	0	0	1	0
chairlady	0	0	0	0
chair	0	0	2	9

Table 4 shows the frequencies of terms beginning with *chair-* (along with plural forms). Here it is clear to see that *chairman* has always been the most popular choice, although there is a rise in the gender-neutral *chair* in the 2006 data (concordance analyses were used to disregard cases of *chairs* which referred to objects that could be sat on). There were two cases of *chairmen* that referred to women (both in the 1991 corpus), again suggesting that this is a rare usage and unlikely to be widely adopted. While there is hope for the gender-neutral *chair* replacing *chairman*, finding a gender neutral term to replace *spokesman* is more problematic. The term *spokes* already exists as an unrelated plural word, while any suffix ending in *-person* may face resistance from some speakers as the suffix sounds earnestly and off-puttingly ‘politically correct’. Around the end of the 1980s, there was a backlash against a range of phenomena that was disparagingly termed ‘political correctness’. Cameron (1995) notes that, ‘PC now has such negative connotations for so many people that the mere invocation of the phrase can move those so labelled to elaborate disclaimers or reduce them to silence’ (1995: 123).

The suffix *-person* is rare in all four corpora, occurring mainly in *spokesperson*. Interestingly, in the 1990s FLOB data, there are single references to *superperson* and *policeperson*. Both are used in ways that are mocking or disapproving of ‘politically correct’ language. One is a humorous ‘politically correct’ version of the children’s story *Noddy in Toyland*, where a character is (somewhat redundantly) described as a *woman policeperson*. In this story, Noddy pours red paint over a teddy bear (for wearing fur), is arrested as a racist for using the word *golly* (as an exclamation) and then telephones a child-abuse helpline to accuse his friend Big Ears of sexually abusing him. Some contextualising information is probably helpful here. In the 1980s, Enid Blyton (1897–1968), the author of *Noddy in Toyland* was accused of using sexism and racism in her children’s books, and they were banned from a number of libraries or rewritten with the offending language removed (see Rudd 2000). The parody of Noddy in the FLOB corpus was taken from the humorous political magazine *Private Eye*. Politically correct people are constructed in this story as militant, misinformed

witch-hunters, suggesting that the writer of the story disapproved of ‘political correctness’.

The other use of *-person* in FLOB is a news article about the television game-show *The Krypton Factor*, where the presenter is derided for his ‘careful’ use of language in calling the winner a ‘superperson’. These two examples are indicative of the ‘PC backlash’ of this period, where certain gender sensitive terms were more likely to be used in order to be criticised or lampooned. There are two non-ironic cases of *craftsperson* in the BE06 corpus (both from the same file), which suggests that the suffix *-person* has not grown in popularity, but is also not subjected to the same level of joking that it had at the height of the PC backlash.

Table 5: Frequent adjectives referring to *man* and *woman*

	BLOB	LOB	FLOB	BE06
man	young 59 old 46 good 19 plain 9 small 8 big 8 tall 8 married 8 great 8 black 7 dead 7	young 63 little 44 old 37 good 18 white 14 big 12 right 10 reasonable 10 great 10 dead 8 married 6 primitive 6 best 6 better 5 new 5 tall 5 nice 5 rich 5 small 5	young 35 old 25 good 18 new 10 big 10 black 9 white 9 little 8 tall 6 poor 6 handsome 5 dead 5 dark 5 nice 5	young 27 old 18 older 11 black 9 new 8 big 7 great 6 best 6 real 6 small 6 married 5 right 5 white 5
woman	young 13 old 10 little 8	old 18 young 16 married 6 grey 5 older 5 little 5	young 17 old 16 older 5	old 20 young 15 fat 6 married 6

A further question about the terms *man* and *woman* involves the sorts of adjectives that are used in relation to them. Table 5 shows frequent adjectives that are used to refer to these words. Because *man* is more frequent than *woman*, there are more words overall that describe men in the corpus. However, it should be noted that there is not a great deal of difference in this table as the majority of frequent adjectives tend to reference age, size and marital status for both *man*

and *woman*. It is perhaps worth pointing out that men are described as *big* and *tall*, whereas *little* features more in the female part of the table (especially in the earlier corpora). There also seem to be a number of positive terms for men (*great, good, nice, handsome, right, reasonable, best* and *better*). Table 5 does not reveal very much about change over time however.

A different approach, inspired by Pearce (2008) was to consider adjectives which described only males or only females. This is shown in Table 6, along with adjectives which are common to both. This table reveals more interesting differences, particularly over time. For example, it is notable in the 1931 corpus that men are described in ways that suggest they are less than (stereotypically) attractive: *dilapidated, ferret-like, plain, pot-bellied, dirty-faced, suet-pudding-faced, thick-set, shabby* and *spectacled*. On the other hand, in the 2006 corpus, men are described as *pretty, hunky, handsome, dapper, Gap-clad* and *fashion conscious*. Clearly, men can be now represented in terms of caring about or looking after their appearance. For women, in the 1931 corpus there is a list of personality traits which are somewhat negative: *vain, obedient, grasping, docile, mad, possessive, talkative, quiet*. These adjectives either describe women as unstable or pet-like. However, in the 2006 corpus, there are some adjectives describing women with positive personality traits: *assertive, carefree, determined, extraordinary, wise, intelligent, remarkable* and *formidable*.

Not all adjectives suggest diachronic change though; some are more indicative of a stagnancy of representation. For example, in all four time periods there are words which collectively refer to men as powerful. Such words include *celebrated, distinguished, noted, rich, wealthy, grand, famous, top, self-made, cultivated, sophisticated* and *dignified*. The words themselves may be different in each time period, but the semantic concept they refer to is similar. Such words are hardly ever used to describe women. There are no references to a 'great woman' or an 'influential woman', not even in the 2006 corpus. Instead, one stable way that women are represented across the four corpora is to do with desirability (*beautiful, pretty, good-looking, fine, desirable, lovely, fresh-complexioned, sensuous, attractive, luscious, voluptuous, sultry, delightful* etc). Again, the individual words may change, but the category remains the same. As shown above, there is evidence that men are starting to be referred to with 'attractiveness' words, although it is of interest that there are no references to *sultry, luscious* or *voluptuous* men. Instead, when men are described as attractive, it tends to be with words like *virile* or *hunky*. Related to this are a set of words across all the time periods which describe men as being physically strong and competent: *able, hale, fastest, healthy, energetic, powerful, strong, mighty, tough*. These words do not tend to be used to describe women, suggesting that

to an extent, men and women continue to be referred to in stereotypical ways. Two other interesting points about the word *man* are worth making. The term *family man* occurs in all four corpora, although there is no equivalent *family woman*, perhaps due to societal assumptions that all women are ‘family women’ by default. Additionally, there are still references to the generic *man* in scientific and historic contexts. An example from the BE06 Corpus is illustrative: ‘The Celts were the second wave of invaders to follow Neolithic man to Britain’ – but presumably not Neolithic woman. However, the generic term *mankind* has fallen in usage, occurring in the 1931, 1961, 1991 and 2006 data 12, 17, 1 and 3 times respectively. On the other hand, the frequency profile for the gender neutral term *humankind* is 0, 0, 1 and 6.

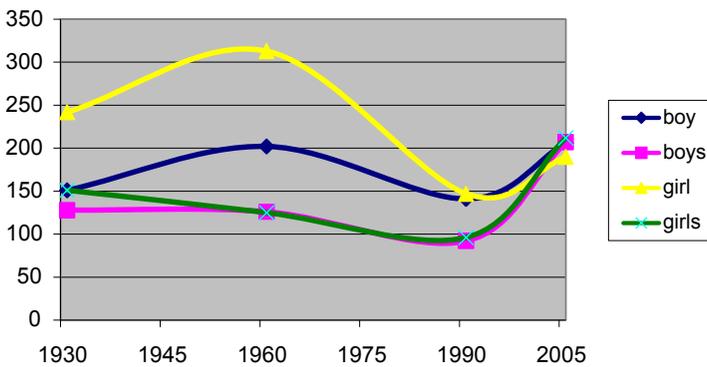


Figure 5: Frequencies of BOY and GIRL

Interestingly, the terms BOY and GIRL show a similar pattern of convergence to MAN and WOMAN (see Figure 5). BOY and GIRL are almost identical in frequency in the 2006 corpus data. Does equal frequency mean equal representation though? Concordance lines of these terms were examined in order to determine the extent to which they were actually used to refer to adults (this was a similar approach to that taken by Sigley and Holmes 2002). In the 1991 FLOB Corpus Sigley and Holmes found that only 3.3% of citations of BOY referred to adults, while this figure was 23.3% for GIRL. I found that references to both BOY and GIRL as adults had increased sharply in the BE06 data, although there were still more cases of GIRL as an adult (52% vs. 28%). Are such uses necessarily sexist though? Some examples from the BE06 data show differences in uses of *boy* and *girl*.

Table 6: All adjectives referring to man and woman

	only for man	only for woman	both
BE06	action, angry, ambitious, aristocratic, bearded, average, best, big, bigger, bird-like, blind, black-browed, bright, career, Gap-clad, compact, fashion-conscious, craggy, cyber, dangerous, dapper, deaf, decent, dignified, disabled, dull, dullest, educated, eerie, energetic, estimable, evil, family, feckless, footsore, free, gay, gentle, good, good-looking, great, gregarious, dark-haired, holy, honorable, hunky, huge, hurt, ideal, injured, lazy, little, lively, loneliest, lonely, murdered, naked, natural, Neolithic, nice, normal, older, open, ordinary, outback, polite, portly, private, quiet, real, religious, right, Renaissance, rugby-playing, powerful, primal, safe, sane, schoolmasterly, shy, secretive, serious, shambling, short, shrewd, silly-looking, slightly-built, sophisticated, straight, strict, strong, stronger, stupid, suburban, tabloid, tattooed, threatening, transparent, true, trusting, unconscious, unknown, untypical, wealthy, wee, wiry, working, warm, weary, wrinkled	ageing, Asian, assertive, bad, beautiful, busy, carefree, determined, easy, English, foreign, formidable, good, copper-haired, frazzled, flicky-haired, fun-loving, grey-haired, harassed-looking, hermit, large, local, moody, poor, pregnant, plump, remarkable, silly, silver-haired, single, sultry, tragic, under-dressed, unusual, Welsh, white, younger	middle-aged, black, blonde, different, elderly, extraordinary, fat, fine, grizzled, grumpy, handsome, hard, intelligent, married, modern, mystery, new, old, passionate, pretty, small, tall, weak, wise, young
FLOB	adventurous, affectionate, aloof, austere, average, babyish, best, big, black-clothed, bluff, bright, cautious, charming, civilized, committee, conservative, conventional-looking, convicted, cynical, dark, darling, dead, decent, despairing, dignified, earnest, excitable, famous, family, fancy, fat, fine-looking, friendly, gangly, generous, good-looking, gentle, great, guilty, grey-haired, gruff-voiced, handsome, happy, happier, hearty, horrible, ignorant, individual, inexpert, influential, innocent, insignificant, jokey, leading, lone, lonely, long-haired, lovely, lucky, macho, nice, one-armed, open-minded, ordinary, outstanding, natural, neighbourly, nice, perceptive, portly, rational, reckless, red-faced, repellent, respectable, retiring, richest, self-made, senile, sensible, silent, silly, smartly-dressed, stocky, tail-coated, thin, thin-faced, unadventurous, unselfish, understanding, untidy, well-bred, white, white-faced, wild, wicked, working, yellow-skinned	aggressive, attractive, beautiful, brown-skinned, busy, career, childless, colourful, depressed, enterprising, filthiest, genuine, hard, home-loving, impassive, independent, irrelevant, irritating, isolated, knowledgeable, large, luscious, married, middle-aged, odd, older, patient, perfect, petite, pleasing, pregnant, red-haired, scrappiest, sensuous, sexiest, shabby, single, slender, small, smart, stimulating, swarthy, Swedish, terrified, thin-lipped, ugliest, unattractive, upmarket, well-built, well-known, voluptuous, witty, wronged, worthless, youngist	black, elderly, extraordinary, good, jolly, little, local, new, old, poor, strange, tall, working-class, young, younger
LOB	able, ablest, active, arrogant, astute, average, bearded, best, best-known, better, big, big-framed, black-faced, black-haired, blind, blond, brave, broken, bronzed, brownish, burly, callow, calm, civilized, common, cultivated, dark-haired, dependable-looking, dowdy, easy, educated, energetic, fair-haired, familiar, family,	acquisitive, ageing, Australian, beaten, chattering, debauched, desirable, French, fresh-complexioned, frail, frightful, gossiping,	business, charming, dead, elderly, experienced, good, grey-haired, hardworking,

	only for man	only for woman	both
LOB	famous, fashionably-dressed, fat, fine, foolhardy, free, gentle, ginger-haired, gay, grand, great, gregarious, good-looking, grown, handsome, happier, happiest, harassed, homeless, ill-mannered, imaginative, imperious, inhuman, inoffensive, intelligent, intemperate, invisible, kind, large, larned, leading, lean, lighter, likeable, limping, living, local, luckiest, mellowed, middle-aged, mighty, modest, mournful, murdered, narrow-shouldered, nice, nice-looking, odd, old-maidish, older, ordinary, overbearing, pale, pale-faced, pink, plain, pleasing, political, practical, precocious, prehistoric, primitive, probing, professional, prudent, puissant, real, reasonable, remarkable, right, ruthless, sandy-haired, sane, severe-looking, sharp-eyed, short, shortish, sick, sincere, slender, slightly-built, small, shrunken, solitary, special, spectacled, steady, stern, strong, thin, top, toughest, troublesome, ugly, unflurried, unfortunate, uniformed, unreliable, untrustworthy, useful, weary, vivitized, wee, well-balanced, western, white-haired, wizened, working, worried, wounded, wrong, youngish	handsomest, hesitant, knowing, lone, lovely, mean, mean-minded, no-good, obese, older, plump, prettiest, raddled, robust, sensible, significant, single, smart-looking, strong-willed, talked-of, wilful	healthy, honest, little, married, naked, old, poor, rich, skilled, tall, unknown, wealthy, white, young, younger
BLOB	able, acid, active, adorable, agile, astute, attractive, bad, best, best-loved, better, big, bigger, bitter, brave, black-bearded, brilliant, broken, business, busy, celebrated, changed, charitable, Christian, clever, coloured, commonplace, competent, coy, curious, dead, deaf, dear, delicate, desperate, determined, different, dilapidated, dirty, dirty-faced, disappointed, disillusioned, distinguished, doomed, drunken, dull, elder, excellent, exemplary, family, fastest, fat, fine-hearted, ferret-like, free, gaunt, glittering, god-fearing, good, grave, great, greatest, hail-fellow-well-met, hairy-faced, hard-headed, hale, hardworking, heavy, heavily-built, highly-trained, honest, horrible, hospitable, humble, hungry, imaginative, impersonal, individual, industrial, innocent, intelligent, kindly, large, leaden, leading, lean, light-haired, long-lived, lonely, lovely, living, medical, medicine, melancholy, murdered, mystery, natural, Neolithic, nervy, nervous, new, noted, older, painstaking, pale, patient, personable, pink, plain, poorer, popular, practical, precise, pot-bellied, primitive, professional, reasonable, red-haired, reliable, reserved, respectably-dressed, rich, richer, self-conscious, shabby, shell-shocked, short, sick, silly, solemn-looking, spectacled, splendid, strong, suet-pudding-faced, sharp-featured, slowest, small, sociable, successful, stupid, talented, tall, tallest, thick-set, tired, tough, touchy, unconscious, unemployed, unhappy, uniformed, universal, urbane, versatile, villainous-looking, virile, wealthy, well-built, well-dressed, well-known, white, wiry, wonderful, working, working class, younger	beautiful, British, broad-bosomed, broad-hipped, deceased, deep, distracted, docile, dream, dying, earnest, elderly, experienced, faded, fascinating, feminine, funny, grasping, ideal, live, mad, middle-aged, modern, obedient, possessive, pretty, proud, sensible, shaken, shrewd, still, soft, strange, talkative, unfortunate, unsmiling, unselfconscious, vain, white-faced	average, dangerous, decent, fine, gentle, good-looking, happiest, little, married, mere, old, poor, quiet, real, stout, thin, wise, young

- (1) ...bare-chested snarling **boy** in blue jeans
- (2) The stereotypical bad **boy** was good looking and sexy
- (3) Nick was a mother's **boy**
- (4) He deployed his pretty-**boy** public school voice
- (5) Barbara's Dad's barrow-**boy** charm
- (6) ...had gone to sea as a cabin **boy**
- (7) A City **boy** which means he's paid well
- (8) The mediaeval custom of appointing a **boy** bishop
- (9) A city analyst who fell in love with a call **girl**
- (10) Heather, dubbed Mucca over her porn and vice **girl** past
- (11) She was that rarest of creatures: a stunning **girl** with a nice uncomplicated attitude to sex
- (12) The future of the British monarchy rests solely on the influence of a blonde party **girl** named Bubble
- (13) Cleaner, cleverer... and faithfully Daddy's **girl** to the end

In the above examples, *boy* could be used to refer to a male who was viewed as rather dangerous and therefore attractive (examples 1 and 2), but it can also indicate a male who is viewed as immature or feminine (examples 3 and 4). *Boy* can also refer to a range of occupations, some which are probably not well-paid (*cabin boy*, *barrow boy*), while others would have better prospects (*city boy*, *boy bishop*). On the other hand quite a significant number of cases of *girl* refer to prostitution (examples 9 and 10) or sex in some way (line 11). Also, *girl* is sometimes used to denote triviality: in example 12 it is viewed as ironic that the future of the British monarchy rests of the influence of a 'blonde party girl'. Additionally, it is worth noting that *Daddy's girl* (line 13) does not seem to have the same negative connotations as *mother's boy* (line 3). So while frequencies of *boy* and *girl* seem to be equalising, there are still qualitative differences in representation, which again are suggestive of gender stereotypes.

The final set of words that I wish to consider are those which refer to titles (*Mr*, *Mrs*, *Miss* and *Ms*). Gendered titles are of particular interest to researchers of language and gender in English-speaking countries because of the inbuilt inequality in the labelling system. Males are not forced to reveal their marital status with *Mr*. However, females are required to reveal their marital status by choosing between *Mrs* or *Miss*. The third option *Ms* was conceived

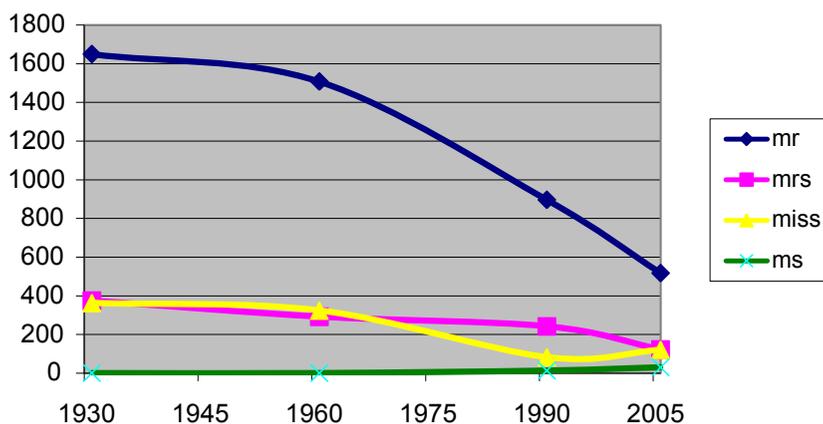


Figure 6: Frequencies of terms of address

in the mid-twentieth century as an equivalent to *Mr*. Pauwels (2003:566), who has researched use of titles in corpora of speech from Australian radio programs, parliamentary debates and academic lectures writes that ‘...*Ms* has been added as a new option besides *Mrs* and *Miss* with the latter titles unlikely to become obsolescent in the near future’. However, *Ms* does not seem to have provided a straightforward resolution. There is sometimes confusion over how it should be pronounced and Schwartz (2006) notes that some people connect it with being divorced or being a lesbian. To what extent has *Ms* being taken up as a strategy? Figure 6 shows frequencies of terms of address. An interesting point to make about this figure is the very high usage of *Mr* in comparison to the female titles. However, it is even more interesting to note that *Mr* appears to have declined over time (especially since 1961). The most frequent two female terms *Mrs* and *Miss* have also decreased (although they were not particularly common in the 1931 data). There appears to have been a small increase in the use of *Ms*, although this term is still very rare. It never occurred in the 1931 and 1961 corpora. There were only nine cases of *Ms* in the 1991 corpus and 30 cases in 2006. In terms of proportions, *Ms* was used 2.7% of the time when people wanted to use a female term of address in 1991, whereas this figure was 10.9% in 2006. So while *Ms* is increasing, it still has a long way to go before it breaks through as the favoured female term of address. However, a more heartening message (for proponents of gender equality) is that rather than writers utilising a new word, they seem to be in the process of abandoning the gender titles system altogether. If the trends in the chart continue for the next 20 to 30 years, all gender marked titles in

English would become very rare, making the question raised in the title of this paper: will *Ms* ever be as frequent as *Mr*, somewhat redundant. A better question might be: will *Mr* ever be as infrequent as *Ms*?

This finding raises another question, why are most of the gendered terms of address reducing over time? There may be a number of different reasons. First, perhaps due to increased awareness of the inequality of the system, people have simply declined to participate, refusing to use any gendered terms of address. A second reason might be that people are not getting married as much as they used to. Figures obtained about the Office of National Statistics⁹ about the General Marriage Rate in the UK show that there has been a steady decline in the numbers of married people since 1980 (which was the earliest figure I was able to obtain). While marriage has decreased, divorce in the UK has increased. According to the Office of National Statistics there were 6 divorces per 1,000 people in 1971, while this figure was 12.2 in 2006. So while people are not getting married as much, they are divorcing more. This could explain why *Mrs* is declining in usage – there simply are not as many women who this word applies to. It may explain the slight increases in *Miss* and *Ms* between 1991 and 2006.

However, the decline in marriage cannot explain the decline in *Mr* as this term of address can refer to married or unmarried men. Perhaps a more probable reason for the decline in titles could be to do with the increasing personalisation of British culture, reflected in language use. This is also argued by Leech (2002) who notes that there have been strong declines in strong modal verbs like *should*, *ought* and *must* between the early 1960s and the early 1990s. Mair (2006) uses phenomena described by Fairclough (1992, 1995) such as democratization, technologization and informalization of public discourse in order to both describe and explain the changes that seem to be happening to British (and American) language use in recent decades. Mair (1997) also coins a related term: colloquialization, the tendency for written English in the twentieth century to move closer towards spoken norms. Therefore, rather than referring to someone as *Mr Smith*, they might be more likely to be addressed in a more personal and emotionally involved way, as *William Smith*, or even an affectionate shortened version like *Bill* (the theory also explains why the informal *cop* is now more popular than the more formal *policeman* or *police officer*).

It could be argued then, that the decline in titles is indicative of several trends: a move towards non-sexist language, a move towards more informal, equal and colloquial ways of addressing people and a (slight) reflection of the decrease of marriage in society.

Conclusion

What can be concluded from the analysis of gendered terms? First, in quantitative terms, male bias is still in existence – males are referred to more often than females. However, this difference seems to be reducing, particularly due to decreases in male terms (notably *he* and *Mr*). There have been smaller increases in female terms, while frequencies of nouns referring to males and females seem to be converging. There is some evidence to suggest that the unequal titles system will be resolved, not by the uptake of *Ms*, but by the decline of the entire system. In defence of *Ms*, however, it should be noted that although the term still does not appear to have been widely taken up, it was perhaps a useful word in that its existence helped to raise consciousness about the problems with the system and may have led to avoidance of other terms.

Some gender-neutral alternatives are being taken up though: *police officer* and *cop* are now more popular than *policeman*, although *spokesman* and *chairman* are still much more common than the gender neutral terms *spokesperson* and *chair/chairperson*. The success of *police officer* could be due to efforts that are being made within the police force to work towards gender equality. *Spokes-* and *chair-people* do not seem to constitute organised professions in the same way, rather, they are roles which people can take on as part of other occupations. There does not appear to be an organisation for women spokespeople or women chairs in the same way that there is an organisation for women in the police, which advocates gender equality.

Another possible reason why *police officer* is rising in popularity is that the word *officer* appears to be more specific and thus appropriate than the generic suffix *-person*. Similarly, I would predict that a term like *fire fighter* would be more successful at replacing *fireman* (although as noted earlier, the frequencies of *fire-* in the corpus data examined are too small to warrant detailed discussion). If *chairman* is to be challenged, it looks as if the existing word *chair* is likely to represent a more successful alternative to *chairperson*. The (admittedly small amount of) evidence from the 1991 data suggests that the *-person* suffix was one strategy that was singled out by opponents of ‘political correctness’ for derision. This may have resulted in this suffix now having too many negative connotations to be successfully taken up, at least in the short term. A solution in keeping with adapting an existing word could be to abandon *spokes-* completely and use another word such as *representative*. In the BE06 there are already nine cases of *representative* being used in similar ways to *spokesperson*.

The patterns in the corpus data suggest a hierarchy of language change in relation to gender equality. First, people seem to be more easily persuaded to stop using a sexist or biased term (such as *Mr*). Second, if a new term must be used in place of an old one, then one which sounds naturalistic (such as

police officer), and is based on existing words or word combinations (such as *chair*) is more likely to be successful. Third, the invention of a completely new term (such as *Ms* or the *-person* suffix) is likely to be met with suspicion and resistance. Terms which are problematic to say (as opposed to write) such as *Ms* (which I have heard pronounced as /mɪz/, /mʊz/ or /mʊs/) or *s/he* are also unlikely to cross over into wider usage, particularly as written language seems to be (increasingly) influenced by spoken language. I would predict that the alliteration in *fire fighter* would also contribute towards its potential success.

The qualitative analysis found some promising trends towards equality in gender representation. Men are more likely to be represented as physically or sexually attractive or caring about their appearance (a kind of equality, even if such objectification and evaluation is not necessarily a good thing). The stereotyping of women as submissive or gossipy seems to be declining somewhat. However, it is still rather surprising to see generic uses of *prehistoric man* and *industrial man* in contemporary academic writing. While this may be a small issue, I would argue that it erases women from history and the practice should be curbed. Related to this is the absence of terms like *great woman* and *influential woman* in any of the corpora studied. Nor do we seem comfortable with representing women as physically strong and able. These findings suggest that efforts could be made to ensure that women are written about in these more positive ways. These qualities are clearly not the preserve of men, and while the lack of terms like *spokeswoman* might reflect a social reality in that not many women take on this role (relative to men), I would argue that there are many cases of strong or influential women in existence. They simply do not seem to be referred to in such ways.

As a coda, reassuringly, the terms *feminism* and *feminist(s)* are continuing to increase over time. Collectively these words occurred three times in 1931, never in 1961, 23 times in 1991 and 59¹⁰ times in 2006. Far from declining, the concept of feminism appears to be doing well.

This paper shows the value of using corpus methods in order to investigate change in the frequency and context of use of gender-marked language over time. While corpus linguistics is yet to make a significant impact on gender and language research, it is hoped that this situation is about to change. A recent book entitled *Gender and Language Research Methodologies* (Harrington et al. 2008), contains three chapters on corpus approaches. With more corpora being made available online, incorporating easy-to-use search facilities, it is now a good time for gender and language researchers to begin to explore these large, untapped resources.

About the author

Paul Baker is a Senior Lecturer at the Department of Linguistics and English Language, Lancaster University, UK.

Notes

- 1 A lemma is the canonical or citation form of a set of inflected forms, and are written in small capitals. So MAN is equal to *man* and *men*.
- 2 See <http://cqpweb.lancs.ac.uk/>.
- 3 See <http://corpus.byu.edu/>.
- 4 I am grateful to one of the reviewers of this paper, who suggested that the word *copper* be examined as a term for the police.
- 5 A Spearman correlation coefficient is a measure of correlation which is based on rank order of items rather than their actual frequency. In the case of comparing different pairs of corpora, the rankings of the most frequent 20 words were used to calculate the Spearman correlation coefficient (see Oakes 2009). For each pair, a score of between 1 and -1 is obtained, where 1 indicates perfect correlation, 0 means no correlation at all and -1 is perfect negative correlation. The higher the score, the more similar the rankings of the top 20 words in the two corpora being compared.
- 6 Interestingly, the main cause of the difference of BE06 to the other corpora, is due to the words *he* and *his* being less frequent in BE06 when compared against the other corpora.
- 7 The word *hers* is very infrequent, so it was combined with *her*.
- 8 The word *they* can also be used as a gender-inclusive pronoun. However, it is difficult to determine how often it is used in this way as the approximately 12,000 cases of *they* across the four corpora would need to be examined qualitatively to determine which ones are gender-inclusive. After examining a sample consisting of a few hundred concordance lines of *they*, this strategy was found to be rare, and also difficult to ascertain, even with a close reading of the context. For these reasons, *they* has not been included in Table 1.
- 9 <http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=9593>.
- 10 Of these 59 cases in the BE06, there are two references to *post-feminism* and two uses of *anti-feminist*. There is one use of *post-feminist* in FLOB.

References

- Baker, Paul (2005) *Public Discourses of Gay Men*. London: Routledge.
- Baker, Paul (2008) *Sexed Texts: Language, Gender and Sexuality*. London: Equinox.
- Baker, Paul (2009) The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics* 14(3): 312–337.
- Biber, Douglas (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas., Johansson, Stig, Leech, Geoffrey, Conrad, Susan and Finegan, Edward (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Brown, Jennifer, Hegarty, Peter and O'Neill, Darragh (2006) Playing with numbers: A discussion paper on positive discrimination as a means of achieving gender equality in the police service in England and Wales. Unpublished paper. University of Surrey. www.bawp.org/assets/file/Playing%20with%20Numbers%20Extract.doc.
- Cameron, Deborah (1995) *Verbal Hygiene*. London: Routledge.
- Fairclough, Norman (1989) *Language and Power*. London: Longman.
- Fairclough, Norman (1992) *Discourse and Social Change*. Cambridge: Polity Press.
- Fairclough, Norman (1994) Conversationalization of public discourse and the authority of the consumer. In Russell Keat, Nigel Whiteley and Nicholas Abercombie (eds) *The Authority of the Consumer* 253–268. London: Routledge.
- Fairclough, Norman (1995) *Critical Discourse Analysis: The Critical Study of Language*. London: Longman.
- Francis, Winthrop Nelson and Kučera, Henry (1979) *Brown Corpus Manual, Revised Version*. Brown University. Accessed from <http://icame.uib.no/brown/bcm.html>.
- Harrington, Kate, Litosseliti, Lia, Sauntson, Helen and Sunderland, Jane (eds) (2008) *Gender and Language Research Methodologies*. London: Palgrave Macmillan.
- Hartman, Pat and Judd, Elliot (1975) Sexism and TESOL Materials. *TESOL Quarterly*. 12(4): 383–393.
- Hunston, Susan (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kennedy, Graeme (1998) *An Introduction to Corpus Linguistics*. London: Longman.
- Kilgarriff, Adam (2001) Comparing corpora. *International Journal of Corpus Linguistics* 6(1): 97–133.
- Kilgarriff, Adam and Tugwell, David (2002) Sketching words. In Marie-Hélène Corréard (ed.) *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins* 125–137. Grenoble: EURALEX.
- Kjellmer, Goeran (1986) 'The lesser man': Observations on the role of women in modern English writings. In Jan Arts and Willem Meijs (eds) *Corpus Linguistics II* 163–176. Amsterdam: Rodopi.
- Lakoff, Robin (1975) *Language and Woman's Place*. New York: Harper and Row.

- Leech, Geoffrey (2002) Recent grammatical change in English: Data, description, theory. In Karin Aijmer and Bengt Altenberg (eds) *Proceedings of the 2002 ICAME Conference* 61–81. Gothenburg.
- Mair, Christian (1997) Parallel corpora: A real-time approach to the study of language change in progress. In Magnus Ljung (ed.) *Corpus-Based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)* 195–209. Amsterdam: Rodopi.
- Mair, Christian (2006) *Twentieth Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.
- McEnery, Tony and Wilson, Andrew (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony, Xiao, Richard and Tono, Yukio (2006) *Corpus-Based Language Studies: An Advanced Resource Book*. London: Routledge.
- Pearce, Michael (2008) Investigating the collocational behaviour of MAN and WOMAN in the British National Corpus using Sketch Engine. *Corpora* 3(1): 1–29.
- Oakes, Michael (2009) Corpus linguistics and language variation. In Paul Baker (ed.) *Contemporary Approaches to Corpus Linguistics* 159–183. London: Continuum.
- Pauwels, Anne (2003) Linguistic sexism and feminist linguistic activism. In Janet Holmes and Miriam Meyerhoff (eds) *The Handbook of Language and Gender* 550–572. Oxford: Blackwell.
- Rayson, Paul, Leech, Geoffrey and Hodges, Mary (1997) Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2(1): 133–152.
- Romaine, Suzanne (2001) A corpus-based view of gender in British and American English. In Marlis Hellinger and Hadumod Bußmann (eds) *Gender across Languages* Vol. 1 153–175. Amsterdam/Philadelphia: John Benjamins.
- Rudd, David (2000) *Enid Blyton and the Mystery of Children's Literature*. London: Macmillan.
- Schmid, Hans-Jörg (2003) Do men and women really live in different cultures? Evidence from the BNC. In Andrew Wilson, Paul Rayson and Tony McEnery (eds) *Corpus Linguistics by the Lune. Łódź Studies in Language* 8 185–221. Frankfurt: Peter Lang.
- Schwarz, Juliane (2006) 'Non-sexist Language' at the Beginning of the 21st Century: Interpretative Repertoires and Evaluation in the Metalinguistic Accounts of Focus Group Participants Representing Differences in Age and Academic Discipline. PhD thesis. Lancaster University.
- Sigley, Robert and Holmes, Janet (2002) Looking at girls in corpora of English. *Journal of English Linguistics* 30(2): 138–157.
- Sunderland, Jane (1996) Gender in the EFL classroom. In Tricia Hedge and Norman Whitney (eds) *Power, Pedagogy and Practice*. Oxford: Oxford University Press.
- Sunderland, Jane (2004) *Gendered Discourses*. London: Palgrave.
- Tognini-Bonelli, Elena (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.