Robbie Love*

Swearing in informal spoken English: 1990s–2010s

Abstract: This paper investigates changes in swearing usage in informal speech using large-scale corpus data, comparing the occurrence and social distribution of swear words in two corpora of informal spoken British English: the demographically-sampled part of the Spoken British National Corpus 1994 (BNC1994) and the Spoken British National Corpus 2014 (BNC2014); the compilation of the latter has facilitated large-scale, diachronic analyses of authentic spoken data on a scale which has, until now, not been possible. A form and frequency analysis of a set of 16 'pure' swear word lemma forms is presented. The findings reveal that swearing occurrence is significantly lower in the Spoken BNC2014 but still within a comparable range to previous studies. Furthermore, FUCK is found to overtake BLOODY as the most popular swear word lemma. Finally, the social distribution of swearing across gender and age groups generally supports the findings of previous research: males still swear more than females, and swearing still peaks in the twenties and declines thereafter. However, the distribution of swearing according to socio-economic status is found to be more complex than expected in the 2010s and requires further investigation. This paper also reflects on some of the methodological challenges associated with making comparisons between the two corpora.

1 Introduction

Swearing is well-established as an area of linguistic enquiry, driven by researchers' curiosity about the "inherent variability and […] subjectivity" (Beers Fägersten 2012) of vocabulary that has the potential to offend and abuse, and yet entertain and create humour, in equal measure. Researchers in corpus linguistics have

---

*Corresponding author: Robbie Love, School of Social Sciences and Humanities, Aston University, Aston Triangle, Birmingham, B4 7ET, UK, E-mail: r.love@aston.ac.uk. https://orcid.org/0000-0002-7212-1165

shown increasing interest in swearing over the last two decades, a noteworthy example being McEnery (2006), who conducted a thorough and unprecedented investigation of swearing in informal spoken English, using the demographically-sampled component of the (first) British National Corpus (BNC1994), which was compiled in the early 1990s (BNC Consortium 2007).

More recently, a second version of the Spoken British National Corpus (BNC2014) – compiled in the mid-2010s – was released, with the express intention of facilitating diachronic comparison of the two datasets (Love et al. 2017). This is the impetus for this paper, as it provides the opportunity to explore what has happened to swearing in casual British English conversation since the 1990s. This article reports on a corpus-based analysis of a set of 16 swear words, comparing their frequency at two diachronic sampling points: the 1990s and 2010s.

I explore the following research questions:

RQ1. Do speakers swear more or less in the 2010s compared to the 1990s?

RQ2. Which swear words are most common in the 2010s, and are these different from the 1990s?

RQ3. What is the variation in the use of swear words according to gender, age and socio-economic status in the 2010s, and is this different from the 1990s?

These questions aim to establish a bird's eye view of swear word frequency over time, focussing on notable differences between the 1990s and 2010s, and offering valuable insights into the occurrence of swearing in the corpora. By using a limited set of 16 swear word lemmata, I am able to use form as an easily-accessible point of direct comparison across time, gaining a perspective on short-term change and variation in swearing occurrence in spoken English which has, until now, not been possible.

The article is structured as follows. In Section 2, I present a literature review, which interrogates the ways in which swearing has been defined in the literature, before coming to the working definition used in this study. Section 3 presents the data and methodology used in this paper and discusses some of the methodological challenges associated with comparing the datasets. Section 4 presents the findings and offers interpretive discussion with reflection on previous research. The conclusion (Section 5) summarises the study and points the way to further research on this topic.

2 Literature review

2.1 What 'counts' as swearing?

Swearing – typically considered to be the use of words which "have the potential to be offensive, inappropriate, objectionable, or unacceptable in any given social

context" (Beers Fägersten 2012: 3) – is "a rich emotional, psychological, and sociocultural phenomenon" (Jay 2009a: 153). Although swearing and societal discourse around swearing have existed for centuries, it has only started to become a prominent subject of research in linguistics (e.g. Andersson and Trudgill 1992; Cheshire 1982; Lakoff 1975; Ljung 2011; McEnery 2006; Partridge 1947), psycholinguistics (e.g. Jay et al. 2008), neurolinguistics (e.g. van Lancker and Cummings 1999), history (e.g. Hughes 1991; Montagu 1967) and other disciplines since the mid-20th century.

Despite such attention in the literature, swearing is notoriously difficult to define, as the human experience of swearing is incredibly subjective. This has implications for the identification, retrieval and analysis of swearing in corpora, as a decision needs to be made about what 'counts' as swearing for the purposes of the research. Even the basic notion that swearing must have some relationship with (potential) offensiveness (e.g. Beers Fägersten 2012) is problematic, since – especially in informal, spontaneous interaction – swearing can also perform functions like expressing humour, creating social bonds and constructing identity (Stapleton 2010). Jay (2009a: 155) discusses the use of swearing to achieve "positive social outcomes", e.g. through humour, sex talk or in-group slang. This is especially true for what Jay calls 'conversational swearing'; Jay's view is that "there is no evidence of harm from fleeting expletives or from conversational or cathartic swearing" (2009b: 93). Therefore, rather than thinking of swearing solely in terms of offensiveness, it can be considered to express "strong emotions or attitudes" and refer to "something taboo or stigmatised" (Andersson and Trudgill 2007: 195); the major taboo areas from which many swear words are derived have been observed to be sex (sexual taboos), bodily functions (excretory/scatological taboos) and religion (profanity) (Stapleton 2010).

These views are generally accepted across the literature. However, there is a division in the literature that has important implications for the identification of swear words. This concerns the question of whether or not the usage of a swear word 'counts' as swearing if the usage is literal, rather than figurative. Some scholars firmly exclude literal usage. For example, another of Andersson and Trudgill's (2007: 195) criteria is that swearing is "not to be interpreted literally". Similarly, in his cross-linguistic study into the "shape, use and manifestations" of swearing, Ljung (2011: viii) puts forward a typology that excludes literal uses of swear words. He asserts that swear words are exclusively emotive in meaning (rather than referential), to the extent that "taboo words with literal meaning cannot be regarded as swearing" (Ljung 2011: 12). The reason given is that swear words, when used in their literal sense (e.g. *He fucked him*), can be replaced by non-taboo near-synonyms (e.g. *He bonked him*), but that the same word, when used in a non-literal sense (e.g. *fuck you*), cannot be replaced by the same set of

synonyms (i.e. **bonk you* is not a suitable replacement). However, counter-examples (such as *screw*, which works in both contexts) weaken the replacement test argument. Others who express a similar view to Andersson and Trudgill (2007) include Lutzky and Kehoe (2015: 167), who "do not regard literal uses of taboo words as swearing (e.g. the word *shit* being used with reference to the excretory system)", the reason being that literal uses are said not to "express emotions".

However, the exclusion of literal usage appears to be the minority view in the literature, with most scholars opting to include literal usage of swear words as legitimate cases of swearing. McEnery (2006: 25) takes an inclusive approach, regarding swearing as one of several types of the wider phenomenon known as 'bad language', the other types being animal terms of abuse, sexist terms of abuse, intellect-based terms of abuse, racist terms of abuse and homophobic terms of abuse. Crucially, he includes the literal usage of swear words in his study. Others in favour of including literal usage include Singleton (2009), who, according to Dynel (2012: 28), claims that, among other criteria, words must "display literal and non-literal senses" in order to be classified as "vulgar". Furthermore, Dynel herself, paraphrasing Hughes (1991), states that swear words "need not convey their basic literal semantic meaning" (Dynel 2012: 28) to be used abusively or to express emotion, for example. Similarly, Drummond (2020) takes what he calls "a common-sense approach", "in which someone saying *'I fucked him'*, at a family meal for example, would most likely be seen as swearing" (p. 3).

Another argument in favour of including literal usage of swear words derives from evidence about the arousing autonomic properties of these words. The strong social conditioning around swearing is such that swear words are more psychologically arousing than other words (Janschewitz 2008), even when encountered as isolated words. Consequently, speakers are so-conditioned that swear words are inherently more memorable than other words (Jay et al. 2008) – again, when encountered as isolated words. Bowers and Pleydell-Pearce (2011) analysed electrodermal activity in participants reading aloud the words *fuck* and *cunt*, and their euphemistic equivalents *f-word* and *c-word*, and found that "people find it more stressful to say aloud a swear word than its corresponding euphemism" (Bowers and Pleydell-Pearce 2011: 4). Therefore, it is difficult to accept the view that swear words (e.g. *shit*) somehow do not trigger such autonomic or cognitive responses when used literally (e.g. to refer to the act of defecating), and yet do trigger psychological arousal when used emotively (e.g. as an interjection).

Thus, in this paper, I adopt an inclusive approach and focus on the forms of swear words, seeking to explore the occurrence of swear word forms regardless of their function. To summarise, swearing can be viewed as a type of so-called bad language, which, when used literally, relates to taboo topics (typically sex, bodily

functions and religion), but can also be used figuratively to perform a range of functions, including abuse, humour and expression of emotion.

2.2 Corpus research into swearing in spoken English

A formal approach to the retrieval and analysis of the occurrence of swear words lends itself to a corpus approach. Corpus linguistics – the analysis of "some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions" (McEnery and Hardie 2012: 1) – has been shown to facilitate investigations of swearing using a range of datasets. Several studies have made use of the British National Corpus 1994 (BNC Consortium 2007) – specifically, its spoken, demographically-sampled component (Spoken BNC1994DS – where 'DS' stands for 'demographically-sampled'), which comprises informal conversation; the metadata associated with the speakers (categories such as gender, age and socio-economic status) have allowed for the analysis of the social distribution of swearing in British English speech. Rayson et al. (1997) identified the most frequent lexical items within different demographic categories of the Spoken BNC1994DS. They found that swear word forms *fucking* and *fuck* were among those significantly more likely to be produced by males, under-35s and those from social classes C2/DE – 'skilled manual workers', 'semi-skilled and unskilled manual workers' and 'state pensioners, casual and lowest grade workers, unemployed with state benefits only' (National Readership Survey 2020).[1]

McEnery et al. (1999, 2000) analysed bad language words ('BLWs', which include swear words and terms of abuse) in the demographically-sampled component of the BNC1994 (the Spoken BNC1994DS), performing a quantitative analysis of the distribution of the BLWs across the sociolinguistic categories of gender, age and socio-economic status. This work led to McEnery's (2006) in-depth analysis of bad language in a bespoke dataset derived from the Spoken BNC1994DS. McEnery (2006) noted that swearing is a "marker of distinction in English" (p. 24) and that "the use or lack of use of BLWs is a fault line along which age, sex and social class may be differentiated" (p. 50). He (2006: 30) found that males draw "typically from a stronger set of words than females", while swearing strength tends to decrease with rising age as well as socio-economic status. In

---

[1] Speakers in the BNC1994 are grouped according to occupation using a classification system known as Social Grade. This comprises six classifications: A, B, C1, C2, D and E. This system has come to be assumed by users of the corpus as an ordinal proxy for socio-economic status (social class), with A being the highest class and E being the lowest. For more information, see http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/ (accessed July 2021).

terms of age, he found that adolescents and young adults are more likely to use bad language words (p. 38) and, in terms of socio-economic status, he reported that "class relates to BLW use in ways in which we might expect (frequency of usage being inverse to height of social class)" (p. 44). McEnery's (2006) study looked at a range of 'bad language' beyond swear words and was one of the first empirical investigations into the phenomenon of swearing in informal spoken English. Subsequently, these findings were echoed in Schweinberger's (2018) study of swearing in Irish English, using the private dialogue section of the Irish component of the International Corpus of English (ICE) (Kallen and Kirk 2008). He found that young adult speakers used swear words most frequently, and that "men used a significantly higher mean rate of swear words compared to women" (p. 11). This shows that there is a precedent for this type of research, and the Spoken BNC2014, with its rich speaker metadata, is well-placed to facilitate the analysis of swearing in comparison to the Spoken BNC1994DS.

Other studies have focussed on particular swear words or investigated particular speech contexts. McEnery and Xiao (2004) published a study entirely devoted to *fuck* and its occurrence in the Spoken BNC1994. They found that males were approximately three times as likely to produce this swear word as females. Detailed analysis of its social distribution revealed that, while young people (especially teenagers) used *fuck* the most, the 35–44 age category had an "unexpectedly low propensity" (McEnery and Xiao 2004: 241) for using this word when compared to the 45–59 age category, which had a higher frequency. They offered the hypothesis those parents of young children, who were likely to populate the 35–44 age categories, may be less likely to say *fuck* than other adults who do not live with children (or those whose children have grown up, and so are more likely to populate a higher age category). Finally, the C1 group ('supervisory, clerical and junior managerial, administrative and professional', National Readership Survey 2020) was found to use *fuck* significantly less than not only C2/DE but also AB ('higher managerial, administrative and professional' and 'intermediate managerial, administrative and professional', National Readership Survey 2020). McEnery and Xiao (2004: 244) speculated that this dip is an example of the members of C1 attempting to "appear closer to what they perceive to be the norms of AB speech".

Focussing just on teenage talk, Stenström (2006) found that *fuck* is the most commonly produced swear word in the Bergen Corpus of London Teenage Language (COLT, Stenström et al. 2002, which is part of the Spoken BNC1994), and that it occurs more than twice as often in boys' speech than girls'. Drange et al. (2014) compared COLT to comparable corpora from Spanish and Norwegian to study 'swearing by mother' (e.g. *motherfucker*) among teenage speakers of the three languages, finding that this is most common in Spanish and least common in

Norwegian. Most recently, Drummond (2020) studied a small corpus of teenage talk gathered from learning centres in north-west England, finding no significant gender difference in the rate of swearing, but much higher overall levels of swearing compared to teenage speakers in the BNC1994 and BNC2014.

Corpus research into swearing makes a contribution to discussions about the typical social distributions of swearing. In terms of gender, the research generally appears to support the long-observed suggestion in the literature that swearing is a characteristic feature of men's language, and that women tend to avoid swearing (see Stapleton 2003 for discussion of this). In addition to social conditioning, males are said to be genetically predisposed to produce stronger swear words more than females due to evolutionary intergroup aggression among males (Güvendir 2015). However, some evidence has emerged recently which suggests there may be some exceptions to this observation. Gauthier (2012) studied perceptions of swearing among L1 English informants. He observed that males over the age of 25 believe they use *fuck* more than females, but that the opposite is true for young adults. Aijmer (2018) investigated intensifiers in the Spoken British National Corpora, with a focus on *fucking*. She reported that *fucking* in the Spoken BNC2014 is "used primarily by young female speakers to mark their linguistic independence" (p. 91).

With regards to age, the corpus research corresponds with other observations that it is young adults who are prone to swearing most frequently (for discussion, see Beers Fägersten 2012: 11–12). However, the case of socio-economic status appears to be more complex, with some (e.g. Johnson and Fine 1985, cited by Beers Fägersten 2012: 52) previously claiming that swearing can be a signal of power and superiority, and others (e.g. Stapleton 2010) discussing the stereotypical association between swearing and "lower levels of education and/or socioeconomic standing" (p. 291). Alongside gender, these social categories are of interest because of the availability of comparable metadata in the Spoken BNC datasets, making it possible to investigate the occurrence of swearing across gender, age and socio-economic groups between the 1990s and 2010s.

3 Data and method

3.1 The data used in this study

For this study, I used the spoken demographically-sampled component of the British National Corpus 1994 (BNC Consortium 2007; Leech 1993; hereafter Spoken BNC1994DS) and the spoken component of the British National Corpus 2014 (Love et al. 2017). The Spoken BNC1994DS contains 5,014,655 tokens across 153 texts, and the Spoken BNC2014 contains 11,422,617 tokens across 1,251 texts. Both corpora

represent informal spoken British English as spoken mostly in England, the former in the early 1990s and the latter in the mid-2010s; the representativeness of both corpora is discussed in detail by Love (2020).

The corpora were accessed using two tools. The first tool, Lancaster University's CQPweb (Corpus Query Processor) interface (Hardie 2012), was used for the analysis of swear word frequency at the whole-corpus level (RQs 1 and 2). For the comparison of swear word occurrence among social groups (RQ3), I used the sociolinguistically-balanced subsets of the corpora, accessible via BNClab (Brezina et al. 2018), due to the established limitations of using an aggregate approach for the comparison of socially-defined subcorpora (Brezina and Meyerhoff 2014). The subsets in BNClab are designed to be roughly comparable in terms of the gender, age, socio-economic status and region of 250 speakers from each of the corpora. Thus, the subsets comprise smaller 'cores' of the original corpora, comprising 2,714,337 and 5,938,032 tokens, respectively, from the Spoken BNC1994DS and the Spoken BNC2014 (Love 2020: 197).

In terms of the comparability of these corpora, it could be argued that, since neither of the Spoken British National Corpora were sampled with the explicit aim of studying swear words, it is difficult to claim that the sampling conditions allowed for a comparable amount of swear word use. However, firstly, it can be assumed that the Spoken BNC1994 facilitated the natural occurrence of swearing, given its surreptitious approach to recording (Crowdy 1993: 260). Secondly, an aim in compiling the Spoken BNC2014 was to facilitate the recording of conversations in a way that minimized intrusiveness in the context of ethics procedures, introduced since the compilation of its predecessor, requiring that speakers provide informed consent before the commencement of recording (Love et al. 2017). Although this does mean that the contexts of recording are not identical, it does not seem to be the case that speakers were inhibited from speaking naturally; a contributor to the Spoken BNC2014, who submitted over a dozen recordings, claimed that "it was surprising how quickly people seemed to forget they were being recorded" (Strawson 2017: 41). This is, of course, anecdotal evidence from only one participant, and further research should be undertaken to explore the possible effects of the observer's paradox (Labov 1972) in the context of spoken corpora; however, McEnery et al. (1999: 51) do state that, while the observer effect may possibly reduce the frequency of bad language, they "see no reason to believe that the patterns of usage for individual [bad language] words are affected by this observer effect".

Another issue relates to the comparison of only two corpora in diachronic research. In this paper, I am interested in change over time; however, I only compare two sampling points – the early 1990s and the mid-2010s. When making such comparisons, the possible outcomes are necessarily limited to three patterns:

an increase between point A and point B, or a decrease, or stability. Without more sampling points, it is difficult to conclude whether an observed change or stasis represents part of a long-existing development or a short-term phenomenon. Creating more sampling points would involve collecting data from further in the past (e.g. the 1930s LOB corpus, Leech and Smith 2005) of sufficient breadth and depth to compile a 'long' historical dataset, and conduct periodisation, which would allow for the tracking of change over time (cf. Nevalainen 2018). An older Spoken BNC, from a comparable point in time, would need to come from the 1970s; although some recordings of casual conversations from this time likely do exist, it is not a reasonable aim to curate a 10-million-word corpus, of a range of UK regions and sociolinguistic groups, from this decade. The further back through the 20th century one looks, the harder this task becomes. Rather, my approach is to compare the two existing corpora, observe findings about difference and stasis, and encourage my research questions to be revisited in the future when more data becomes available. Therefore, I frame the findings presented in this paper with the point in mind that they are not alone directly indicative of language change (or stasis) per se, but rather suggestive of change or stasis.

3.2 Analytical approach

In this study, I analysed a set of 16 swear words, the lemmata of which are listed here:

> ARSE, BASTARD, BITCH, BLOODY, BOLLOCK, BUGGER, COCK, CRAP, CUNT, DICK, FUCK, PISS, SHAG, SHIT, TWAT, WANK

In selecting these swear words, I made several considerations, based on the lack of a consensus within the research community about what 'counts' as swearing, the result of which, as Beers Fägersten (2012: 4) points out, is that "the category of swear words remains open ended". Likewise, Drummond (2020: 6) states that, "[g]iven the vagaries surrounding the definitions of swearing […], it is quite difficult to systematically justify including or excluding particular words". Accepting these limitations, in selecting these swear words, I focussed on words derived from sexual taboos, excretory/scatological taboos and profanity (Stapleton 2010). Also, based on my specific interest in conducting a wholesale analysis of forms, I sought to isolate words that were likely to be almost always used to refer to something taboo (regardless of literal or figurative usage) in casual conversation. For example, *fuck* does not have a common usage that is outside of the context of swearing, so it can be considered a 'pure' swear word. On the other hand, *pussy*, which can be used to refer to the vagina (sexual taboo), can also commonly be used

to refer to a cat, and, in the context of swearing, that usage would skew the data relating to the use of *pussy* as a swear word. The resulting set reflects the subjective nature of swearing, and while some may disagree with these judgements, I endeavoured to isolate an easily-searchable and unambiguous set of the most semantically 'pure' swear words.

I then created search queries for the swear words both in CQPweb and BNClab, while considering the morphological variants of each swear word that should be captured in the search. In CQPweb, I used CQL (corpus query language) syntax to refine the queries for precision, and originally planned to use lemma searches for swear words which were, clearly, morphological headwords (e.g. CRAP). This would have simplified the search queries and ensured that no rare or unexpected morphological variants were omitted from the search (i.e. maximizing recall). However, the automatic lemmatization of the corpus data was not reliable enough to do this; for example, a lemma search for CRAP in the Spoken BNC2014 ([lemma = "crap"%c]) retrieves 545 instances, comprising *crap, crapper, crapped, crappest, crapping* and *craps*, but the non-lemma search ([word = "crap.*"%c]) retrieves 625 matches, including 77 instances of the adjective *crappy*, which were not included by the lemmatization. This observation forced me to abandon searching for lemma forms and adapt the search queries using wildcard characters; the limitations of the lemmatization in the corpora is worthy of further investigation in the future.

In BNClab, I used the simple syntax that is afforded by the tool (Brezina et al. 2018). Since this came after searching for the swear words in the full corpora (in CQPweb), I knew all of the possible forms that could occur in the BNClab subsets, so I was able to search for each individual form (e.g. crap OR crappy OR crapped OR crapper OR craps OR crapola OR crappest OR crapping). A full list of the swear words, their search terms and frequencies in the full corpora (extracted from CQPweb) can be found in the Appendix A.

Once the appropriate search queries were written, I was then able to address the Research Questions using the following procedure.

- For RQ1, I searched in the Spoken BNC1994DS and Spoken BNC2014 in CQPweb (Hardie 2012) for each of the swear words in turn and compared the overall relative frequency of the whole set between the two corpora.[2]
- For RQ2, I investigated which individual swear words differed in relative frequency to the greatest extent between the two corpora.

---

2 Significance (log-likelihood) and effect size (log ratio) was calculated using the Lancaster UCREL (University Centre for Computer Corpus Research on Language) log-likelihood and effect size calculator, which is available at http://ucrel.lancs.ac.uk/llwizard.html (accessed July 2021).

–   For RQ3, I used the sociolinguistically-balanced subsets of the two corpora in
     BNClab (Brezina et al. 2018) to analyse the social distribution of swearing
     according to speaker gender, age and socio-economic status.

4 Findings

4.1 Wholesale frequency analysis

The total frequency of all of the swear words in the full versions of the corpora and
the socially-balanced subsets is presented in Table 1.

   The table shows that swear word usage is significantly lower in the 2010s data
than the 1990s data. Interestingly, the BNClab subsets for both corpora retrieve a
higher relative frequency of these swear words than the full corpora; the 1994 subset
frequency is 24% higher than the BNC1994DS frequency, and the 2014 subset fre-
quency is 8% higher than the Spoken BNC2014 frequency. This shows that the 250
speakers selected for inclusion in each of the BNClab subsets produce swear words at a
higher rate than those speakers who are excluded from the subsets – perhaps because
the socially-balanced BNClab subsets exclude less prolific contributors to the corpora,
potentially inflating the swearing rate for the speakers who remain. Since the inflation
is larger for the 1994 subset (24% compared to 8%), it does amplify the decrease in the
swearing rate, as indicated by the larger effect size for the BNClab data (0.64)
compared to the full corpora (0.47). Although I only speculate about the cause of this
inflation here, it is worth acknowledging the amplifying effect in interpreting the social
distributions of these swear words.

   This finding supports existing research showing that swearing generally takes
up a very small proportion of speaker output; swearing accounts for 0.23 and
0.14% of tokens in the 1990s and 2010s BNClab data respectively. This is compa-
rable to, although slightly lower than, the findings of Jay (2009b: 90), who
reviewed several empirical studies into bad language and reported that swearing
constitutes 0.3–0.7% of speakers' output. One reason for this may be that I have

Table 1: Frequency comparison (words per million) of the 16 *true swear words* in the Spoken
BNC1994DS and the Spoken BNC2014 (full corpora and *BNClab* subsets).

	WPM (Spoken BNC1994DS)	WPM (Spoken BNC2014)	Log-likelihood	Significant? (p<0.05)	Effect size (Log Ratio)
Full corpus	1,822	1,320	574.21	Yes	0.47
BNClab subset	2,265	1,428	681.89	Yes	0.64

analysed a restricted set of swear words, while the studies discussed by Jay (2009b) may have considered larger sets of taboo words. However, Schweinberger (2018), who, similarly, focussed on a small set of swear words (22 in total), reported a rate of 0.16% in a corpus of Irish English informal private dialogues, so rates lower than 0.3% in casual conversation are not unprecedented. Therefore, it would be difficult to claim that the difference in frequency between the corpora is suggestive of some wider decline in the use of swear words among British English speakers; both rates appear to be close to what might be expected. A more parsimonious explanation for the decrease, if it is meaningful, is that the difference in speaker awareness that they were being recorded has had an effect, as discussed by Love (2020); perhaps the Spoken BNC2014 speakers, who were aware of the recordings taking place, were slightly less likely to swear as often as their Spoken BNC1994DS predecessors, who were mostly unaware of the recording process (see also Crowdy 1993).

Figures 1 and 2 show the frequency of the individual swear words in the 1990s and 2010s corpora. Both Figures show that British English swearing appears to centre around three main lemmata: FUCK, SHIT and BLOODY. However, while FUCK has remained stable and SHIT has significantly increased in frequency ($p < 0.05$; log ratio 1.09), BLOODY has significantly decreased in frequency ($p < 0.05$; log ratio 2.34), and has done so to the greatest extent out of the nine swear words that significantly decreased in frequency. The other nine declining swear words are (from highest to lowest effect size):
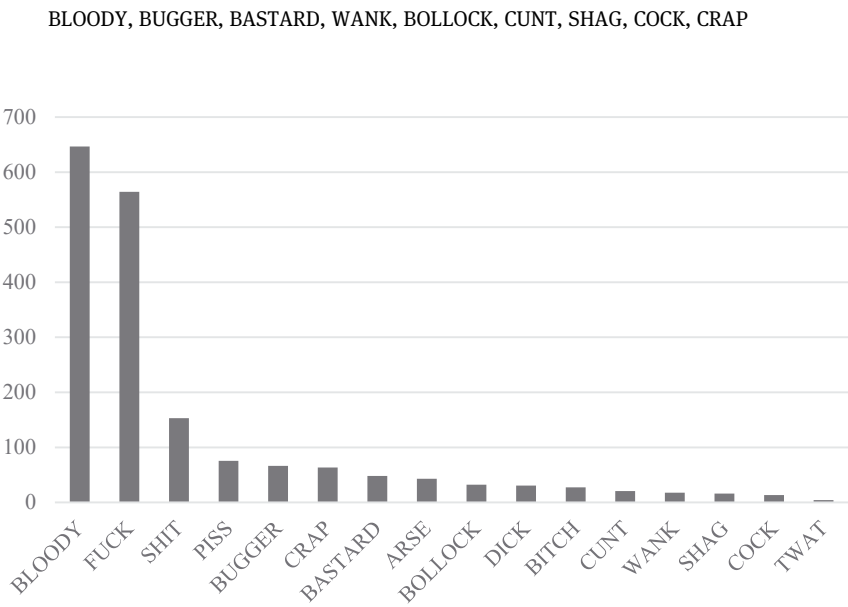
BLOODY, BUGGER, BASTARD, WANK, BOLLOCK, CUNT, SHAG, COCK, CRAP

Figure 1: Relative frequency of swear words (per million) in the Spoken BNC1994DS.
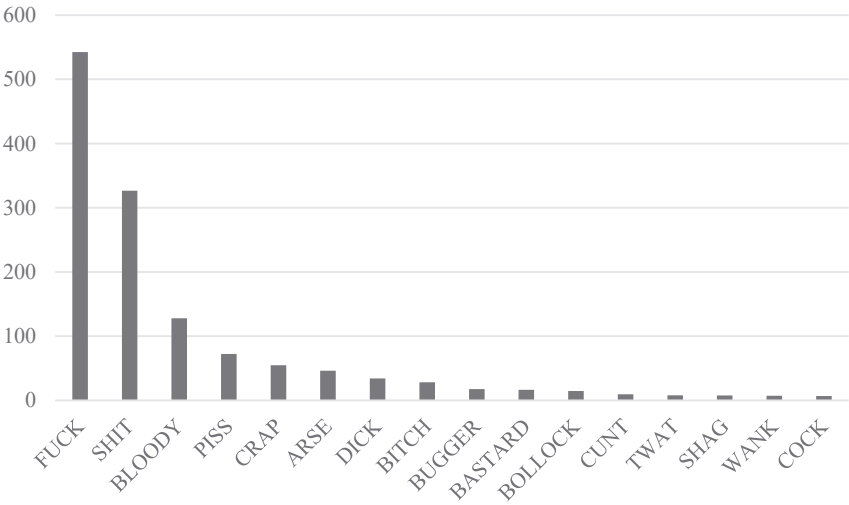
Figure 2: Relative frequency of swear words (per million) in the Spoken BNC2014.

Alongside SHIT, one other swear word – TWAT – increased in frequency significantly (p<0.05; log ratio 0.93), while FUCK, PISS, BITCH, ARSE and DICK do not differ in frequency significantly between the corpora.

The result of these shifts in frequency is that FUCK has replaced BLOODY as the most common swear word. The dominance of FUCK is predicted in the literature; although FUCK is said to have first appeared in English around the year 1,500, its development in the late 1900s is described by Ljung (2011: 71) as "a success story of almost unlikely proportions". It has recently become a highly frequent and productive swear word, despite it being considered one of the strongest bad language words of all (McEnery 2006), and there was already evidence of it being the most common swear word among British teenagers in the 1990s (Stenström 2006). Based on this, it may be the case that FUCK was indeed already the most popular swear word in the 1990s; qualitative work is needed to establish whether the frequency of BLOODY in the 1990s was inflated by its non-taboo usage as a literal adjective (i.e. describing something that may involve blood).

4.2 Social distribution

4.2.1 Gender

Turning to the first of the social categories – gender – Figure 3 shows the distribution of the swear words according to gender in both BNClab subsets. This shows

that male speakers use the swear words more than female speakers in both corpora. Before commenting on the difference between the corpora, the finding that swearing is more than double as frequent among the males (31.65 per ten thousand) than the females (13.56 per ten thousand) in the 1990s data is noteworthy in and of itself, since others have already conducted similar research on the same data and come to a different conclusion. McEnery (2006: 29) is unequivocal: "when all of the words in [his study] are considered, it is equally likely that bad language will be used by a male as by a female". Bearing in mind that his study included a much wider range of 'bad language words', this may indicate that the 'pure' swear words that I have studied are more typical of male speech, but that other bad language vocabulary, including a range of terms of abuse, are more equally distributed. This is useful to know, as this is distinct from McEnery's (2006) observation that males prefer 'stronger' swear words, because the swear words I have analysed represent the full range of strengths on McEnery's (2006) scale of offence. It may be that it is not just the strength, but also the 'purity', of swear words (i.e. lack of polysemy with everyday senses) that distinguishes the genders, with males being more likely to use 'pure' swear words.

In the 2010s data, it is also the case that males produce swear words more so than females (18.15 per ten thousand and 10.81 per ten thousand, respectively), within the context of an overall decline in swearing usage. This is in line with other studies (Beers Fägersten 2012; Schweinberger 2018); it has also been suggested that female swearing is "undoubtedly subject to more social censure" than male swearing (Stapleton 2010: 293) and is therefore likely to be less frequent. However, the reduction from male usage being 2.33 times more frequent than female usage in
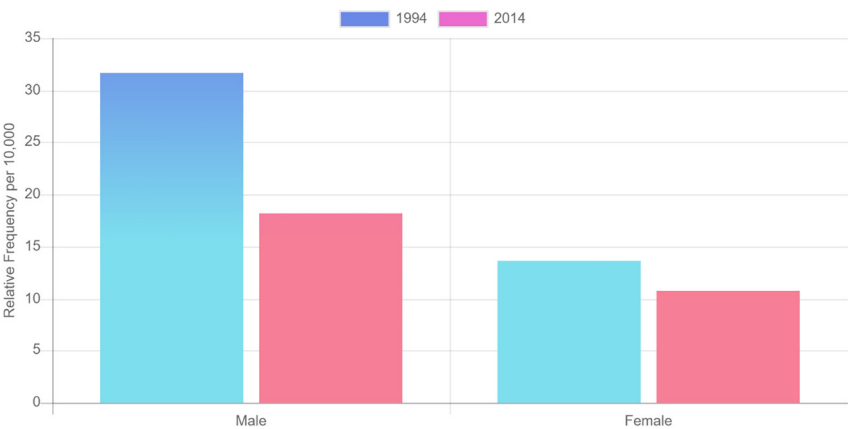

Figure 3: Gender distribution of swearing in the 1994 and 2014 BNClab subsets.

the 1990s to 1.68 times more likely in the 2010s requires further attention. It is known that usage of swearing is highly sensitive to contextual factors, such as the gender of the interlocutor(s), and so further work is needed to divide the data into further meaningful units of comparison.

4.2.2 Age

Figure 4 shows the distribution of swear word frequency according to the age of the speakers in the BNClab subsets. Swearing frequency is distributed similarly in both corpora; overall, frequency decreases with age (as indicated by the trendlines); it rises throughout childhood and adolescence and peaks in the twenties, before steadily declining thereafter. At the top end of the age scale, McEnery (2006) hypothesises that the low level of swear word use in speakers over the age of 60 could be attributed to euphemistic replacement terms being used in places where swear words may otherwise be expected (e.g. *oh sugar* instead of *oh shit*). In the 2010s subset, it is not surprising to observe that lower frequencies are observed among older speakers (and children), perhaps for the reason that McEnery (2006) suggests.

There is also the sharp decline in swear word usage among the late 20s/early 30s, which appears in both subsets (this is also observed in Irish English by Schweinberger 2018). When discussing age grading (specifically in the case of
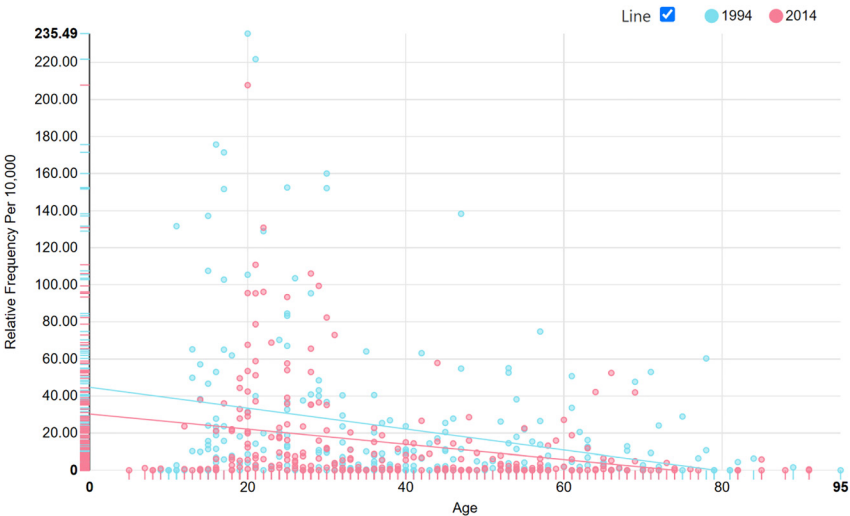
Figure 4: Age distribution of swearing in the 1994 and 2014 BNClab subsets.

FUCK), McEnery and Xiao (2004) suggest that swearing drops rather steeply at middle age, due to this being the typical age at which adults become parents and are said to be less likely to swear due to often being in the presence of children. What is interesting is that, in the 2010s subset, the negative correlation between age and swear word frequency is weaker than in the 1990s subset. In other words, swearing is better distributed across age groups. Perhaps the decline in usage in the middle age is less severe; it could be the case that, nowadays, people of typical parental age simply swear more around their children. There is probably not enough data of that type to assess in the Spoken BNC2014, but, on intuition alone, this suggestion seems unlikely. The more probable answer, in my view, lies in a demonstrable change in UK society which has been in progress for the last few decades: the steady increase in the average age of parents. According to the UK Office for National Statistics, the average age of parents in England and Wales rose by almost four years between the 1970s and 2010s.[3] Therefore, a more horizontal trendline could still be explained by the same hypothesis, so long as we take into account that parents are typically older in the 2010s when compared to the 1990s.

Overall, the negative correlation between age and swearing is entirely expected (Beers Fägersten 2012; McEnery 2006) – the phenomenon of speakers (being perceived as, or otherwise) becoming more conservative with age is by no means unique to the 1990s.

4.2.3 Socio-economic status

Figure 5 shows the distribution of swearing frequency according to socio-economic status. It should be noted that the socio-economic status categories used in BNClab are different from those in the corpus metadata for either corpora. It appears that: 'middle class' comprises Social Grade categories A, B and C1; 'working class' comprises categories C2 and D; and 'student' and 'retired' have been separated out of category E.

McEnery (2006: 44) reported a negative correlation between swearing and socio-economic status. The distribution in Figure 5 conforms to that pattern fairly well for the 1990s data; working class usage (40.53 per ten thousand words) is much higher than that of the middle class (8.48) and retired groups (9.68), and even the student group (25.85). However, the picture is very different in the 2010s subset. Swearing among working class speakers is much lower (11.81), only 29.1% of the 1990s frequency. This, combined with a slight rise in usage in the middle

---

3 Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ livebirths/bulletins/birthsbyparentscharacteristicsinenglandandwales/2015 (accessed November 2020).
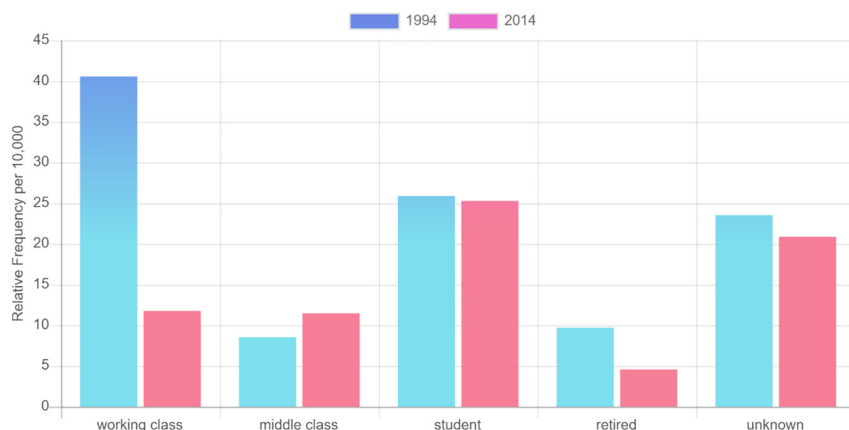
Figure 5: Socio-economic status distribution of swearing in the 1994 and 2014 BNClab subsets.

class (11.53), renders their frequencies almost equal in the 2010s subset, while the student group is responsible for the most swearing (25.27).

For the working class frequency to be so low in the 2010s data is unexpected and difficult to explain based on the wealth of research on the use of swearing among this group (see Stapleton 2010 for discussion). One explanation may be that the age grading effect is dictating the pattern that is observed in the 2010s data. Students typically represent the exact age band which peaks in swearing usage in Figure 4 (early 20s); in the 2010s BNClab data, the students' mean age is 19. Conversely, the middle class and retired groups in BNClab are older (mean age 41 and 70 respectively) and, as discussed, older people are less likely to swear anyway. What requires further investigation is why the same cannot be said of the 1990s data. It could be the case that the amplifying effect of the speaker selection in the 1990s BNClab subset is condensed around a small number of working-class speakers, which may distort the frequency of swear word usage in the 1990s working class group.

5 Conclusion

In this paper, I have investigated the occurrence of swearing in informal spoken British English, using national corpora from the 1990s and 2010s. My aim was to explore how the amount and social distribution of a set of 'pure' swear words may have changed between those decades. I adopted a view of swearing that is inclusive of both literal and figurative usage and accepts all usage of swear word

forms, regardless of function. Corpus linguistics lends itself to this approach, as I could retrieve and analyse the swear words with an exclusive focus on form and frequency.

Returning to the Research Questions, I have found that, overall, swear word usage is significantly lower in the Spoken BNC2014 compared to the Spoken BNC1994DS (RQ1); however, swearing is generally a low-frequency feature of conversational vocabulary, and the rates in both datasets are close to expected levels as predicted by previous research. Secondly, it was found that FUCK appears to have overtaken BLOODY as the most popular swear word in casual conversation, and that SHIT is one of only two swear words to have increased in usage over time, despite the overall decline in swearing (RQ2). Thirdly, an examination of the social distribution of my set of 16 swear words suggests that (a) it may be not only the strength but also the 'purity' of swear words that distinguishes male and female swearing occurrence; (b) the age grading effect on swearing is less severe in the 2010s, perhaps caused in part by a societal increase in parental age; and (c) there is a substantial difference in terms of socio-economic status, which requires further investigation of the corpus data and metadata categorisation.

Further research should interrogate these findings from a functional perspective, taking into account contextual factors such as the relative perceived strength, the pragmatic function and the conversational context of the swearing. Even though the Spoken BNC2014 comprises only one genre of speech (casual conversation), it has been shown that, within the corpus, there is a diverse variety of functional conversational units (Egbert et al. this special issue; see also Love et al. 2019), which may well encourage or discourage swearing. There is also the effect of gender, whereby "most people swear more around listeners of the same gender than in mixed crowds" (Jay and Janschewitz 2008: 274), which should be explored with the Spoken BNC2014 and other contemporary spoken corpora, as McEnery (2006) did with the Spoken BNC1994. Finally, further research could compare the use of swearing in spoken English to other modes of communication, including writing. The compilation of the Spoken BNC2014 has facilitated large-scale, diachronic analyses of authentic spoken data on a scale which has, until now, not been possible. In doing so, it breathes new life into the continued use of the original British National Corpus, which Geoffrey Leech played a large role in compiling.

Appendix A: Search terms and CQPweb frequency data for the swear words included in this study

Swear word lemma	CQP syntax	BNClab syntax	Spoken BNC1994DS		Spoken BNC2014	
			Hits	WPM	Hits	WPM
ARSE	[word = "arse|arses|arsed| arsehole.*|ass|asses| assed|asshat.*| asshole.*"%c]	Arse OR arsehole OR arsed OR ass OR arseholes OR arses OR arseholed OR asshole OR assholes OR asses OR assed	216	43.07	528	46.22
BASTARD	[word = "bastard.*"%c]	Bastard OR bastards	241	48.06	187	16.37
BITCH	[word = "bitch|bitches| biatch|biatches"%c]	Bitch OR bitches OR biatch	137	27.32	321	28.10
BLOODY	[word = "bloody"%c]	Bloody	3,243	646.70	1,459	127.73
BOLLOCK	[word = "bollock.*"%c]	Bollocks OR bollock OR bollocking OR bollocksed OR bollock-less OR bol-lockings OR bol-locked OR bollocksing OR bollocksy	161	32.11	167	14.62
BUGGER	[word = "bugger.*"%c]	Bugger OR buggers OR buggered OR buggery OR bugger-ing OR buggeration	333	66.41	201	17.60
COCK	[word = "cock|cocks"%c]	Cock OR cocks	67	13.36	78	6.83
CRAP	[word = "crap.*"%c]	Crap OR crappy OR crapped OR crapper OR craps OR crapola OR crappest OR crapping	318	63.41	625	54.72
CUNT	[word = "cunt.*"%c]	Cunt OR cunts OR cuntish OR cunting	103	20.54	109	9.54
DICK	[word = "dick|dicks|dick-wad|dickhead.*"%c]	Dick OR dickhead OR dicks OR dickheads	153	30.51	391	34.23
FUCK	[word = ".*fuck.*"%c]	Fucking OR fuck OR fucked OR fucker OR	2,830	564.35	6,195	542.35

(continued)

Swear word lemma	CQP syntax	BNClab syntax	Spoken BNC1994DS		Spoken BNC2014	
			Hits	WPM	Hits	WPM
		fucks OR mother-fuckers OR fuckers OR motherfucker OR mother-fucker OR fluctu-fucking-ation OR fuckhead OR mother-fucking OR motherfuckers OR fuckity OR fuckered OR fuckwit OR fuck-wits OR fucked-up OR fucktard				
PISS	[word = "piss.*"%c]	Pissed OR piss OR pissing OR pisses OR pissy OR pisser OR piss-up OR piss-take OR piss-poor OR pisshead	378	75.38	826	72.31
SHAG	[word = "shag.*"%c]	shag OR shagging OR shagged OR shaga-lagging OR shag-a-lagging OR shagalagged	80	15.95	88	7.70
SHIT	[word = ".*shit.*"%c]	Shit OR shitty OR shitting OR shite OR shits OR shitted OR apeshit OR shitler OR shitter OR shit-arse OR shit-hot OR shit-house OR shittiest OR shittily OR shitloads OR shitload OR shit-test OR birdshit OR shithole OR batshit OR shitbag OR shit-faced OR apeshit OR dipshit OR gobshite OR shit-pants OR shitfaced OR shitless OR pigshit OR shitiest	768	153.15	3,729	326.46

(continued)

Swear word lemma	CQP syntax	BNClab syntax	Spoken BNC1994DS		Spoken BNC2014	
			Hits	WPM	Hits	WPM
TWAT	[word = "twat.*"%c]	OR shiting OR shit-munchers OR shit-ness OR shitstorm OR shitters OR bullshit Twat OR twats OR twattish OR twatted OR twatty	21	4.19	91	7.97
WANK	[word = "wank.*"%c]	Wanker OR wank OR wankers OR wanking OR wanked OR wanks OR wanky OR wankered	89	17.75	82	7.18
Total			9,138	1,822	15,077	1,320

References

Aijmer, Karin. 2018. "That's well bad": Some new intensifiers in spoken British English. In Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014*, 60–95. London: Routledge.

Andersson, Lars & Peter Trudgill. 1992. *Bad language*. London: Penguin.

Andersson, Lars & Peter Trudgill. 2007. Swearing. In Leila Monaghan, Jane Goodman & Jennifer Robinson (eds.), *A cultural approach to interpersonal communication*, 195–199. Oxford: Blackwell.

BNC Consortium. 2007. *The British National Corpus, version 3 (BNC XML Edition)*. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/ (accessed 1 December 2020).

Beers Fägersten, Kristy. 2012. *Who's swearing now? The social aspects of conversational swearing*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Bowers, Jeffrey S. & Christopher W. Pleydell-Pearce. 2011. Swearing, euphemisms, and linguistic relativity. *PloS One* 6(7). e22341.

Brezina, Vaclav, Gablasova Dana & Susan Reichelt. 2018. *BNClab [electronic resource]*. Lancaster University. http://corpora.lancs.ac.uk/bnclab (accessed 1 December 2020).

Brezina, Vaclav & Miriam Meyerhoff. 2014. Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics* 19(1). 1–28.

Cheshire, Jenny. 1982. *Variation in an English Dialect*. Cambridge: Cambridge University Press.

Crowdy, Steve. 1993. Spoken corpus design. *Literary and Linguistic Computing* 8(4). 259–265.

Drange, Eli-Marie Danbolt, Ingrid Kristine Hasund & Anna-Brita Stenström. 2014. "Your mum!": Teenagers' swearing by mother in English, Spanish and Norwegian. *International Journal of Corpus Linguistics* 19(1). 29–59.

Drummond, Rob. 2020. Teenage swearing in the UK. *English World-Wide* 41(1). 59–88. [pre-print].

Dynel, Marta. 2012. Swearing methodologically: The (im)politeness of expletives in anonymous commentaries on YouTube. *Journal of English Studies* 10. 25–50.

Egbert, Jesse, Stacey Wizner, Daniel Keller, Doug Biber, Paul Baker & Tony McEnery. 2021. Identifying and describing functional conversation units in the BNC Spoken 2014. *Text & Talk* 41(5–6). 715–737.

Güvendir, Emre. 2015. Why are males inclined to use strong swear words more than females? An evolutionary explanation based on male intergroup aggressiveness. *Language Sciences* 50. 133–139.

Hardie, Andrew. 2012. CQPweb – Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3). 380–409.

Hughes, Geoffrey. 1991. *Swearing: A social history of foul language, oaths and profanity in English*. London: Blackwell.

Janschewitz, Kristin. 2008. Taboo, emotionally-valenced, and emotionally-neutral word norms. *Behavior Research Methods, Instruments, & Computers* 40. 1065–1074.

Jay, Timothy. 2009a. The utility and ubiquity of taboo words. *Perspectives on Psychological Science* 4. 153–161.

Jay, Timothy. 2009b. Do offensive words harm people? *Psychology, Public Policy, and Law* 15. 81–101.

Jay, Timothy, Catherine Caldwell-Harris & Krista King. 2008. Recalling taboo and nontaboo words. *American Journal of Psychology* 121. 83–103.

Jay, Timothy & Kristin Janschewitz. 2008. The pragmatics of swearing. *Journal of Politeness Research* 4. 267–288.

Kallen, Jeffrey L. & John M. Kirk. 2008. *ICE-Ireland: A user's guide*. Belfast: Cló Ollscoil na Banríona.

Labov, William. 1972. *Language in the inner city*. Philadelphia: University of Pennsylvania Press.

Lakoff, Robin. 1975. *Language and a woman's place*. New York: Harper & Row.

van Lancker, Diana & Jeffrey L. Cummings. 1999. Expletives: Neurolinguistic and neurobehavioural perspectives on swearing. *Brain Research Reviews* 31. 83–104.

Leech, Geoff. 1993. 100 million words of English. *English Today* 9(1). 9–15.

Leech, Geoff & Nick Smith. 2005. Extending the possibilities of corpus-based research in the twentieth century: A prequel to LOB and FLOB. *ICAME Journal* 29. 83–98.

Ljung, Magnus. 2011. *Swearing: A cross-cultural linguistic study*. New York: Palgrave Macmillan.

Love, Robbie. 2020. *Overcoming challenges in corpus construction: The Spoken British National Corpus 2014*. New York: Routledge.

Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3). 319–344.

Love, Robbie, Vaclav Brezina, Tony McEnery, Abi Hawtin, Andrew Hardie & Claire Dembry. 2019. Functional variation in the Spoken BNC2014 and the potential for register analysis. *Register Studies* 1(2). 296–317.

Lutzky, Ursula & Andrew Kehoe. 2015. Your blog is (the) shit: A corpus linguistic approach to the identification of swearing in computer mediated communication. *International Journal of Corpus Linguistics* 21(2). 165–191.

McEnery, Tony. 2006. *Swearing in English: Bad language, purity and power from 1586 to the present*. New York: Routledge.

McEnery, Tony, Paul Baker & Andrew Hardie. 1999. Assessing claims about language use with corpus data – swearing and abuse. In John M. Kirk (ed.), *Corpora galore: Papers from ICAME 1998*, 45–55. Amsterdam: Rodopi.

McEnery, Tony, Paul Baker & Andrew Hardie. 2000. Swearing and abuse in modern British English. In Barbara Lewandowska-Tomaszczyk & Patrick James Melia (eds.), *PALC '99: Practical Applications in Language Corpora*, 37–48. Frankfurt am Main: Peter Lang.

McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, Tony & Richard Xiao. 2004. Swearing in modern British English: the case of FUCK in the BNC. *Language and Literature* 13(3). 235–268.

Montagu, Ashley. 1967. *The anatomy of swearing*. London: Macmillan.

NRS (National Readership Survey). 2020. *Social Grade*. http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/ (accessed 16 November 2020).

Nevalainen, Terttu. 2018. Research methods: Periodization and statistical techniques. In Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (eds.), *Patterns of change in 18th-century English: A sociolinguistic approach*, 61–74. Amsterdam: John Benjamins.

Partridge, Eric. 1947. *Usage and abusage*. London: Hamish Hamilton.

Rayson, Paul, Geoff Leech & Mary Hodges. 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.

Schweinberger, Martin. 2018. Swearing in Irish English – A corpus-based quantitative analysis of the sociolinguistics of swearing. *Lingua* 209. 1–20.

Singleton, David. 2009. Unspeakable words: the taboo fringe of the lexicon. In Marta Dynel (ed.), *Advances in discourse approaches*, 130–146. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Stapleton, Karyn. 2003. Gender and swearing: A community practice. *Women and Language* 26(2). 22–33.

Stapleton, Karyn. 2010. Swearing. In Miriam A. Locher & Sage L. Graham (eds.), *Interpersonal pragmatics*, 289–306. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110214338.2.289.

Stenström, Anna-Brita. 2006. Taboo words in teenage talk: London and Madrid girls' conversations compared. *Spanish in Context* 3(1). 115–38.

Stenström, Anna-Brita, Gisle Andersen & Ingrid Kristine Hasund. 2002. *Trends in Teenage Talk: Corpus compilation, analysis and findings*. Amsterdam: John Benjamins.

Strawson, Harry. 2017. Diary: The British National Corpus. *London Review of Books* 39(6). https://www.lrb.co.uk/v39/n06/harry-strawson/diary (accessed 1 December 2020).

Bionote

Robbie Love
School of Social Sciences and Humanities, Aston University, Birmingham, UK
r.love@aston.ac.uk
https://orcid.org/0000-0002-7212-1165

Robbie Love is a Lecturer in English Language at Aston University, UK, where he is Convenor of the Aston Corpus Linguistics Research Group. His research interests include corpus linguistics, applied linguistics and spoken English. He is author of *Overcoming Challenges in Corpus Construction* (Routledge, 2020) and co-editor of *Corpus Approaches to Contemporary British Speech* (Routledge, 2018). He is a member of the British Association for Applied Linguistics (BAAL) Executive Committee and Communications Officer for the BAAL Corpus Linguistics Special Interest Group.