Language and Literature

http://lal.sagepub.com

Swearing in Modern British English: The Case of Fuck in the BNC

Anthony McEnery and Zhonghua Xiao

Language and Literature 2004; 13; 235
DOI: 10.1177/0963947004044873

The online version of this article can be found at:
http://lal.sagepub.com/cgi/content/abstract/13/3/235

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Poetics and Linguistics Association

ARTICLE

Swearing in modern British English: the case of *fuck* in the BNC

Anthony McEnery and Zhonghua Xiao, *Lancaster University, UK*

Abstract

Swearing is a part of everyday language use. To date it has been infrequently studied, though some recent work on swearing in American English, Australian English and British English has addressed the topic. Nonetheless, there is still no systematic account of swear-words in English. In terms of approaches, swearing has been approached from the points of view of history, lexicography, psycholinguistics and semantics. There have been few studies of swearing based on sociolinguistic variables such as gender, age and social class. Such a study has been difficult in the absence of corpus resources. With the production of the British National Corpus (BNC), a 100,000,000-word balanced corpus of modern British English, such a study became possible. In addition to parts of speech, the corpus is richly annotated with metadata pertaining to demographic features such as age, gender and social class, and textual features such as register, publication medium and domain. While bad language may be related to religion (e.g. *Jesus*, *heaven*, *hell* and *damn*), sex (e.g. *fuck*), racism (e.g. *nigger*), defecation (e.g. *shit*), homophobia (e.g. *queer*) and other matters, we will, in this article, examine only the pattern of uses of *fuck* and its morphological variants, because this is a typical swear-word that occurs frequently in the BNC. This article will build and expand upon the examination of *fuck* by McEnery et al. (2000) by examining the distribution pattern of *fuck* within and across spoken and written registers.

Keywords: *BNC*; *corpus*; fuck; *sociolinguistic variable*; *swear*

1 Introduction

Swearing is a part of everyday language use. To date it has been infrequently studied, though some recent work on swearing in American English (e.g. Jay, 1992), Australian English (e.g. Kidman, 1993) and British English (e.g. McEnery et al., 2000) has addressed the topic. Nonetheless, there is still no systematic account of swear-words in English.[1] In terms of approaches, swearing has been approached from the points of view of history (e.g. Montagu, 1973[1967]; Hughes, 1991), lexicography (Sheidlower, 1995), psycholinguistics (e.g. Jay, 1992) and semantics (Kidman, 1993). There have been few studies of swearing based on sociolinguistic variables such as gender, age and social class.[2] Such a study has been difficult in the absence of corpus resources. With the production of the British National Corpus (BNC), a 100,000,000-word balanced corpus of modern British English,[3] such a study became possible. In addition to parts of speech, the corpus is richly annotated with metadata pertaining to demographic features such as age, gender and social class, and textual features such as register,

Language and Literature Copyright © 2004 SAGE Publications
(London, Thousand Oaks, CA and New Delhi), Vol 13(3): 235–268
DOI: 10.1177/0963947004044873 www.sagepublications.com

publication medium and domain. In this article, we will explore such dimensions of variation in order to discover a general pattern of usage for one word, *fuck*, in modern British English.[4] While bad language may be related to religion (e.g. *Jesus*, *heaven*, *hell* and *damn*), sex (e.g. *fuck* and *cunt*), racism (e.g. *nigger*), defecation (e.g. *shit* and *piss*), homophobia (e.g. *queer*) and other matters, we decided to examine only the distribution pattern of *fuck* and its morphological variants because *fuck* is a typical swear-word that occurs frequently in the BNC. *Fuck* is perhaps 'one of the most interesting and colourful words in the English language today' that can be used to describe pain, pleasure, hatred and even love (Andersson and Trudgill, 1992: 60). As the word becomes more highly charged semantically (see section 5), it has also acquired more grammatical flexibility so that *fuck* 'has altered from being exclusively a verb to every part of speech' (Nurmi, 1997).

This article has two primary goals. Above all, we want to explore one swear-word in detail and, via a corpus-based description of the word and its interaction with a number of register-based and sociolinguistic variables, produce an account of it which allows us to reflect upon claims made about swearing in English in the literature. As such, this article is largely descriptive. However, following from this detailed descriptive work a second goal arises in the form of a question reflecting on the corpus methodology: what are the limitations of the use of corpus data in the study of language? In pursuit of these goals, this article is split into six major sections. Section 2 compares the use of *fuck* in spoken and written language. Section 3 explores the pattern of *fuck* usage in speech. Section 4 explores the pattern of *fuck* usage in writing. Section 5 discusses eight categories of *fuck* usage and section 6 concludes the article.

2 Spoken vs written register

The spoken register is generally more informal than the written register. One of the linguistic indicators of informality is swearing (Collins and Hollo, 2000). In the BNC corpus, the spoken section consists of around 10 percent of the data while the other 90 percent are written texts. Table 1 gives the frequencies of *fuck* used in the spoken and written sections of the BNC.[5] As can be seen from the table, for all the word forms under examination the difference between speech and writing is statistically significant at the level $p < 0.001$. *Fuck* occurs 12 times more frequently in speech than in writing. The greatest contrast is found for *fucking*, which was used nearly 20 times as frequently in the spoken as in the written section of the corpus. While it is not clear why people use *fuck* considerably more in speech than in writing, our speculation is that *fuck* occurs more frequently in informal rather than formal contexts, though the censorship of published written texts is another possible explanation for the relatively lower frequency of *fuck* in writing. In spite of this quantitative difference, different word forms distribute across register in the same descending order: *fucking*, *fuck*,

Table 1 Spoken vs written register

Form	*Register*	*Words*	*RF*	*NF*	*Log-likelihood (LL) score*[6]	*Sig. level*
fuck	Spoken	10,365,464	583	56.24	940.406	<0.001
	Written	89,740,543	795	8.86		
fucked	Spoken	10,365,464	62	5.98	68.066	<0.001
	Written	89,740,543	130	1.45		
fucks	Spoken	10,365,464	10	0.96	12.792	<0.001
	Written	89,740,543	18	0.2		
fucking	Spoken	10,365,464	2164	208.77	6150.587	<0.001
	Written	89,740,543	969	10.8		
fucker(s)	Spoken	10,365,464	25	2.41	28.841	<0.001
	Written	89,740,543	50	0.56		
All forms	Spoken	10,365,464	2844	274.37	6827.547	<0.001
	Written	89,740,543	1962	21.86		

fucked, *fucker(s)* and *fucks*. However, the general difference between spoken and written uses of *fuck* obscures a number of finer differences in usage both within the general discussions of speech and writing and between them. The rest of this article is devoted to identifying these finer distinctions.

3 Variations within spoken English

This section explores the pattern of *fuck* usage in spoken British English using metadata pertaining to the different sociolinguistic variables encoded in the BNC. We will compare demographically sampled and context-governed speech, as well as the types of interaction, and also examine the possible influence of domain, speaker gender, age, social class and education level on the pattern of uses of *fuck*.

3.1 Demographically sampled vs context-governed speech

The BNC corpus contains orthographically transcribed speech collected using two different sampling regimes: demographically determined and context-governed. With regard to the frequency of *fuck*, the two types of speech differ significantly at the level $p < 0.001$. As can be seen in Table 2, demographically sampled speech contains 146 times as many instances of *fuck* as context-governed speech. Some word forms, e.g. *fucks* and *fucker(s)*, are simply non-existent in context-governed speech, even though this category contains nearly one million more tokens than the first type of data.

Surprisingly, the contrast between the two types of speech is even more marked than the contrast between spoken and written registers. While context-

Table 2 Demographically sampled vs context-governed speech

Form	*Type*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
fuck	Demographic	4,211,216	576	136.78	838.609	<0.001
	Context-governed	5,034,707	7	1.39		
fucked	Demographic	4,211,216	61	14.49	86.922	<0.001
	Context-governed	5,034,707	1	0.2		
fucks	Demographic	4,211,216	10	2.37	15.729	<0.001
	Context-governed	5,034,707	0	0		
fucking	Demographic	4,211,216	2149	510.3	3218.681	<0.001
	Context-governed	5,034,707	15	2.98		
fucker(s)	Demographic	4,211,216	25	5.94	39.321	<0.001
	Context-governed	5,034,707	0	0		
All forms	Demographic	4,211,216	2821	669.88	4196.573	<0.001
	Context-governed	5,034,707	23	4.57		

governed speech is indeed more formal than demographically sampled speech (cf. Aston and Burnard, 1998: 31), the difference between the two types of speech cannot be explained by the formal/informal distinction alone: writing is basically more formal than speech, yet the contrast between writing and speech is not as striking as that between the two types of speech considered here. A reasonable explanation for this is that the social contexts from which the context-governed data were sampled militated in favour of considerably fewer forms of *fuck* than in the demographically sampled speech. This explanation will be considered in more detail in sections 3.2 and 5.7.

3.2 Domain

When people talk or write on different subjects, their language use may vary (Collins and Hollo, 2000). The BNC spoken corpus has a context-governed section which attempts to model some of that contextually dependent language use by sampling data from four domains: business, education/informative, public/institutional and leisure.

Table 3 shows that of the four domains, *fuck* occurs most frequently in business. *Fuck* is used 7 times as frequently in business as in education/informative, 9 times as frequently as in public/institutional, and nearly 19 times as

Table 3 Domains of context-governed speech

Domain	*Words*	*RF*	*NF*	*LL score*	*Sig. level*	*LL score*	*Sig. level*
Business	1,310,327	18	13.74	–	–		
Edu. info.	1,028,734	2	1.94			28.114	<0.001
Pub. Inst.	1,327,740	2	1.51	0.898	0.748		
Leisure	1,367,906	1	0.73				

frequently as in leisure. The LL test indicates that the distribution of *fuck* in the domains of education/informative, public/institutional and leisure does not differ signficantly (LL = 0.898, $p = 0.748$), because *fuck* occurs infrequently in all of these domains. The contrast between business and these domains, however, is statistically significant (LL = 28.114, $p < 0.001$). One possible explanation for the high frequency of *fuck* in the business context is that arguments/disputes are common in this domain. It is in just such contexts that swear-words may be used as a rhetorical device.

3.3 Type of interaction

There are two broad types of interaction in the spoken register: dialogue and monologue. In terms of participants, more than one party contributes to a dialogue while only one party contributes to a monologue. In terms of function, a monologue basically gives information whereas a dialogue typically involves exchanging information. While both types of interaction can be spontaneous or scripted/prepared, a dialogue is more likely to be spontaneous than a monologue, because in a spontaneous dialogue the feedback is relatively unpredictable whereas a monologue does not involve verbal feedback. These differences also influence the pattern of usage of *fuck* in modern British English, where the word typically occurs in dialogues, as shown in Table 4.

As can be seen in the column *NF*, when we take all word forms together, *fuck* occurs 150 times more frequently in dialogues than in monologues. This difference is significant at the $p < 0.001$ level. Even when we consider the word forms separately, the differences in the frequencies of *fuck* in the two types of interaction are all statistically significant, though different word forms show different levels of significance. Hence one can conclude that a dialogue is indeed different from a monologue in terms of the usage of the word *fuck*.

Table 4 Types of interaction in speech

Form	*Interaction*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
fuck	Dialogue	7,622,718	579	75.96	204.251	<0.001
	Monologue	1,626,672	2	1.23		
fucked	Dialogue	7,622,718	62	8.13	23.985	<0.001
	Monologue	1,626,672	0	0		
fucks	Dialogue	7,622,718	10	1.31	3.868	<0.049
	Monologue	1,626,672	0	0		
fucking	Dialogue	7,622,718	2147	281.66	805.495	0.001
	Monologue	1,626,672	2	1.23		
fucker(s)	Dialogue	7,622,718	25	3.28	9.671	0.002
	Monologue	1,626,672	0	0		
All forms	Dialogue	7,622,718	2823	370.34	1045.138	<0.001
	Monologue	1,626,672	4	2.46		

3.4 Gender of speaker

Men and women differ in their use of strong language (cf. Lakoff, 1975: 5; Hughes, 1991: 211; Holmes, 1992: 171–6). For example, Stenström (1991) found in the London–Lund spoken corpus that male speakers prefer *hell*-related words like *damn* and *devil* while female speakers show a preference for *heaven*-related words like *heavens* and *gosh*.[7] Consequently we decided to explore the hypothesis that the gender of speakers also influences the frequency of their use of *fuck*.[8] Table 5 compares male and female speakers' use of *fuck*. As can be seen from the normalized frequencies, when all word forms are taken as a whole male speakers use *fuck* more than twice as freqently as female speakers, a difference that is statistically significant at the level of $p < 0.001$. When we consider word forms individually, we find that male speakers use *fucking*, *fuck* and *fucker(s)* significantly more frequently than female speakers. The difference in the frequencies of male and female speakers' use of *fucked* and *fucks* is, however, not statistically significant. One possible explanation is that the two word forms tend to denote the literal meaning of the word (cf. section 5.1).

Table 5 Gender of speaker

Form	*Gender*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
fuck	Male	4,918,075	337	68.52	50.025	<0.001
	Female	3,255,533	106	32.56		
fucked	Male	4,918,075	25	5.08	0.510	0.475
	Female	3,255,533	13	3.99		
fucks	Male	4,918,075	5	1.02	0.386	0.534
	Female	3,255,533	2	0.61		
fucking	Male	4,918,075	1394	283.44	353.624	<0.001
	Female	3,255,533	321	98.6		
fucker(s)	Male	4,918,075	18	3.66	8.967	0.003
	Female	3,255,533	2	0.61		
All forms	Male	4,918,075	1779	361.73	401.668	<0.001
	Female	3,255,533	444	136.38		

Table 6 Proportion and rank of word forms by male and female speakers

Gender	*Form*	*Proportion (%)*	*Rank*
Male	*fucking*	78.36	1
	fuck	18.94	2
	fucked	1.41	3
	fucker(s)	1.01	4
	fucks	0.28	5
Female	*fucking*	72.30	1
	fuck	23.87	2
	fucked	2.93	3
	fucker(s)	0.45	4/5
	fucks	0.45	4/5

On the other hand, while the use of *fuck* differs quantitatively by speaker gender, it does not differ qualitatively (cf. also section 5.3). For both males and females, the rank and proportion of different word forms show a very similar distribution pattern (Table 6). Both genders use *fucking* most frequently, followed by *fuck*. While the proportions of the different word forms may vary slightly by gender, this variation is not statistically significant, as shown in Table 7.

Table 7 Comparison of the normalized frequencies of word forms across gender

Form	*Male*	*Femle*	*LL score*	*Sig. level*
fuck	68.52	32.56		
fucked	5.08	3.99		
fucks	1.02	0.61	4.17	0.35
fucking	283.44	98.6		
fucker(s)	3.66	0.61		

3.5 Age of speaker

Speaker age is another sociolinguistic variable that influences the pattern of *fuck* usage. As Holmes observes:

> The extensive swear word vocabulary which some teenagers use is likely to change over time [...] Though they continue to know these terms, the frequency with which they use them often diminishes, especially as they begin to have children and socialise with others with young families. (1992: 183)

Holmes's hypothesis is supported by our data. Table 8 gives the frequencies of *fuck* for different age groups; it shows that for each of the forms of *fuck*, and for all of the word forms taken together, the difference in the distribution of *fuck* across the different age groups is statistically significant, though the significance level varies by word form, with the most marked contrast being for *fucking*, followed by *fuck*. For all age groups, the most frequently used word form is *fucking*, followed by *fuck*, though the other word forms do not show a predictable pattern.

With respect to age group, young people and teenagers (age groups 15–24 and 25–34) appear to use *fuck* more frequently than people from other age groups (Table 9). While it is not surprising that young people use *fuck* readily, children of the age group 0–14 appear to show an unexpectedly marked propensity to say *fuck* whereas people aged 35–44 demonstrate an unexpectedly low propensity. One plausible reason, in line with Holmes's hypothesis, for the relatively low frequency for age group 35–44 is that parents with children and teenagers around them say *fuck* less than those who have yet to have or do not have children and those whose children have grown up and do not live with them. One might also hypothesize that children under the age of 15 use *fuck* more frequently because they consciously want to behave in a way that they perceive as being more adult. However, on the basis of corpus data alone we cannot evaluate these possible

Table 8 Age of speaker

Form	*Age*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	0–14	460,627	158	343.01		
	15–24	511,858	126	246.16		
fuck	25–34	1,113,709	93	83.50	622.580	<0.001
	35–44	1,066,857	8	7.50		
	45–59	1,605,978	46	28.64		
	60+	1,122,133	3	2.67		
	0–14	460,627	2	4.34		
	15–24	511,858	10	19.54		
fucked	25–34	1,113,709	5	4.49	29.912	<0.001
	35–44	1.066,857	1	0.94		
	45–59	1,605,978	2	1.25		
	60+	1,122,133	0	0		
	0–14	460,627	3	6.51		
	15–24	511,858	1	1.95		
fucks	25–34	1,113,709	1	0.90	11.097	0.015
	35–44	1,066,857	0	0		
	45–59	1,605,978	2	1.25		
	60+	1,122,133	0	0		
	0–14	460,627	217	471.10		
	15–24	511,858	638	1246.44		
fucking	25–34	1,113,709	582	522.58	1967.681	<0.001
	35–44	1,066,857	71	66.55		
	45–59	1,605,978	173	107.72		
	60+	1,122,133	18	16.04		
	0–14	460,627	12	26.05		
	15–24	511,858	18	35.17		
fucker(s)	25–34	1,113,709	8	7.18	88.829	<0.001
	35–44	1,066,857	0	0		
	45–59	1,605,978	0	0		
	60+	1,122,133	0	0		
	0–14	460,627	392	851.01		
	15–24	511,858	793	1549.26		
All forms	25–34	1,113,709	689	618.65	2613.071	<0.001
	35–44	1,066,857	80	74.99		
	45–59	1,605,978	223	138.86		
	60+	1,122,133	21	18.71		

explanations, a point returned to later (see section 6). What we can do with the corpus is refine our view of the data: if we cross-tabulate the variables speaker age and gender, a more distinct pattern can be observed (Table 10).

As can be seen from Table 10, except for the age group 60+, the difference between male and female speakers is statistically significant. For all age groups, male speakers say *fuck* more frequently than female speakers. The greatest

Table 9 Frequencies of *fuck* by age group

Age	*NF*	*Rank by NF*
15–24	1549.26	1
0–14	851.01	2
25–34	618.65	3
45–59	138.86	4
35–44	74.99	5
60+	18.71	6

Table 10 Cross-tabulation of speaker age and gender

Age	*Gender*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
0–14	Male	237,530	248	1044.08	21.77	<0.001
	Female	223,092	144	645.47		
15–24	Male	215,310	657	3051.41	558.717	<0.001
	Female	296,548	136	458.61		
25–34	Male	543,791	643	1182.44	645.124	<0.001
	Female	569,709	46	80.74		
35–44	Male	557,551	64	114.79	26.657	<0.001
	Female	509,306	16	31.42		
45–59	Male	531,429	88	165.59	3.93	0.047
	Female	1,072,944	135	125.82		
60+	Male	531,692	14	26.33	3.17	0.84
	Female	590,441	7	11.86		

contrast between male and female speakers is found in young people (age groups 25–34 and 15–24), as reflected by their much greater LL scores.

3.6 Social class of speaker

The BNC corpus classifies speakers into four social classes, namely AB, C1, C2 and DE. In this section, we examine the possible influence of social class on the distribution pattern of *fuck*. Table 11 gives the frequencies of *fuck* used by different social classes. As can be seen from the table, except for the word form *fucks*, the difference in the distribution of all other word forms across social class is statistically significant.[9] As with speaker gender and age, the greatest contrast is for *fucking*, followed by *fuck*, as indicated by their LL scores. The overall frequencies of *fuck* also show that the distinction between social classes is quantitatively significant.

The normalized frequencies for all forms show a clear distinction. People from classes DE and C2 are the most frequent users, followed by AB. Interestingly, AB speakers say *fuck* more than C1 speakers. This is particularly true of people in the age group 60+ (see Table 13). One might speculate that the older people from AB use *fuck* more frequently because they want to flaunt their seniority, while those

Table 11 Speaker social class

Form	*Class*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	AB	696,819	93	133.46		
fuck	C1	427,872	7	16.36	75.494	<0.001
	C2	485,682	45	92.65		
	DE	267,818	55	205.36		
	AB	696,819	18	25.83		
fucked	C1	427,872	0	0	15.993	0.001
	C2	485,682	4	8.24		
	DE	267,818	2	7.47		
	AB	696,819	3	4.31		
fucks	C1	427,872	0	0	1.987	0.583
	C2	485,682	1	2.06		
	DE	267,818	1	3.73		
	AB	696,819	187	268.36		
fucking	C1	427,872	39	91.15	297.527	<0.001
	C2	485,682	305	627.98		
	DE	267,818	198	739.31		
	AB	696,819	1	1.44		
fucker(s)	C1	427,872	0	0	8.087	0.012
	C2	485,682	2	4.12		
	DE	267,818	4	14.94		
	AB	696,819	302	433.4		
All forms	C1	427,872	46	107.51	339.734	<0.001
	C2	485,682	357	735.05		
	DE	267,818	260	970.81		

from C1 show a considerably lower rate of *fuck* because they consciously or unconsciously pay special attention to their linguistic behaviour so as to appear closer to what they perceive to be the norms of AB speech. This observation is further supported by the cross-tabulation of speaker gender and social class on the one hand, and of speaker age and social class on the other, as shown in Tables 12 and 13.

Table 12 shows the result of cross-tabulation of gender and social class. As can

Table 12 Cross-tabulation of speaker gender and social class

Class	*Gender*	*Words*	*RF*	*NF*	*Ll score*	*Sig. level*
AB	Male	266,857	175	655.78	42.934	<0.001
	Female	413,150	127	307.39		
C1	Male	187,946	43	228.79	52.035	<0.001
	Female	239,926	3	12.5		
C2	Male	169,737	348	2050.23	654.976	<0.001
	Female	315,945	9	28.49		
DE	Male	126,512	176	1391.17	64.701	<0.001
	Female	138,247	84	607.61		

Table 13 Cross-tabulation of speaker age and social class

Age	*Class*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	AB	127,228	209	1642.72		
0–14	C1	5,722	0	0	24.550	<0.001
	C2	4,439	1	225.28		
	DE	2	0	0		
	AB	78,210	80	1022.89		
15–24	C1	40,544	1	24.66	99.486	<0.001
	C2	29,072	29	977.52		
	DE	42,303	81	1914.76		
	AB	101,503	0	0		
25–34	C1	55,654	26	467.17	312.701	<0.001
	C2	192,484	317	1646.89		
	DE	23,468	4	170.44		
	AB	81,002	2	24.69		
35–44	C1	201,306	17	84.45	4.813	0.090
	C2	97,480	10	102.59		
	DE	0	0	0		
	AB	132,275	0	0		
45–59	C1	106,972	2	18.7	431.876	<0.001
	C2	84,611	0	0		
	DE	115,857	168	1450.06		
	AB	94,332	7	74.21		
60+	C1	17,674	0	0	7.835	0.023
	C2	77,596	0	0		
	DE	48,244	0	0		

be seen from the table, while the difference between male and female speakers is statistically significant for all social classes, the greatest contrast is found for the class C2. Male and female speakers of the class DE show a much less marked contrast as both sexes from this class use *fuck* very frequently.

In Table 13, we find that except for the age group 35–44, where the frequency of *fuck* is relatively low, the difference between social classes in all other age groups is statistically significant. The age group 35–44 does not show a significant contrast because people of this age group use *fuck* rarely. The greatest difference is found in the age group 45–59, where no uses of *fuck* are found for classes AB and C2 while the class DE uses *fuck* very frequently. However, non-corpus-based research into the relationship between swearing and power is clearly needed to substantiate further the hypothesis that those in authority flaunt their seniority through the use of swear-words.

3.7 Education level of speaker

A belief is that the better educated one is, the less likely one is to use bad

language.[10] The BNC encodes information pertaining to speakers' education level, thus enabling us to test the influence of education on the use of *fuck*. Table 14 shows the overall frequencies of *fuck* across different levels of education.[11] As can be seen, people who left school at 15/16 are the most frequent users of *fuck*. The general pattern of uses of *fuck* is that people who have received less education say *fuck* more frequently. People who left school at 14 or under show an unexpectedly low frequency of uses of *fuck* because people from this group are mostly over 60.[12] Of the nine instances of *fuck* for this group, only two are used by young people aged 15–24 while seven are used by people aged 60 or over.

In terms of word forms, the distinction across education level is quantitative rather than qualitative. For people of all levels of education, *fucking* is the most frequent word form, followed by *fuck* (see Table 15).

Table 14 Speaker education level

Education	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
Left school 15/16	639,039	596	932.57		
Left school 17/18	217, 282	32	147.27	762.703	<0.001
Educ. until 19/over	318,267	16	50.27		
Left school 14/under	378,669	9	23.77		

Table 15 Comparison of normalized frequencies of *fuck* across education level

Education level	*Word form*	*NF*	*Rank*
	fucking	21.13	1
Left school 14/under	*fuck*	2.64	2
	fucked	0	–
	fucker(s)	0	–
	fucks	0	–
Left school 15/16	*fucking*	772.97	1
	fuck	143.95	2
	fucked	6.26	3/4
	fucker(s)	6.26	3/4
	fucks	3.13	5
	fucking	110.46	1
Left school 17/18	*fuck*	36.82	2
	fucked	0	–
	fucker(s)	0	–
	fucks	0	–
Educ. until 19/over	*fucking*	31.42	1
	fuck	12.57	2
	fucked	6.28	3
	fucker(s)	0	–
	fucks	0	–

4 Variations within written English

This section explores the distribution pattern of *fuck* in written British English using metadata pertaining to the different sociolinguistic variables encoded in the BNC. We examine the possible influence of domain, gender and age of author, gender, age and level of audience, and reception status, publication medium and creation date of texts on the pattern of uses of *fuck*.

4.1 Domain

In section 3.2, we found that frequencies of *fuck* vary across domain in spoken English. *Fuck* typically occurs in the business domain. This section examines the distribution of *fuck* in writing. Table 16 compares the nine written domains encoded in the BNC.

Clearly, the distribution of *fuck* by domain in written English is statistically significant. Forms of *fuck* are used most frequently in imaginative writing, probably because texts of this category are primarily fiction, which contains a lot of representations of speech, and are hence to some extent more speech-like in parts. This is followed by the domains of arts and leisure. In contrast, *fuck* occurs rarely in the domains of belief/thought and is non-existent in natural/pure science. This distribution pattern also applies to the individual word forms.

Table 16 Domains of the written section of the BNC

Domain	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
Imaginative	19,664,309	1485	75.52		
Arts	7,014,792	208	29.65		
Leisure	8,991,740	98	10.9		
World affairs	15,243,340	73	4.79		
Commerce/business	6,668,357	29	4.35	2827.945	<0.001
Social science	12,186,378	45	3.69		
Applied science	7,341,375	21	2.86		
Belief/thought	3,035,896	3	0.99		
Natural/pure science	3,746,901	0	0		

Interestingly, while the business domain in spoken English uses *fuck* most frequently (13.74 instances per million words), it is not used markedly frequently in written English, with a normalized frequency of only 4.35 per million words. Conversely, while *fuck* is least likely to be found in the domain of leisure in speech (0.73 occurrences per million words), it occupies a prominent position in the leisure domain in written English (10.9 occurrences per million words), below only the domains of imaginative writing and arts.

4.2 Gender of author

One would hypothesize that gender has similar effect on the pattern of uses of *fuck* in writing to that in speech. This hypothesis is supported by the data.

Table 17 Gender of author

Form	*Gender*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
fuck	Male	31,586,324	486	15.39	28.625	<0.001
	Female	15,497,994	147	9.49		
fucked	Male	31,586,324	78	2.47	7.549	0.007
	Female	15,497,994	20	1.29		
fucks	Male	31,586,324	14	0.44	6.503	0.029
	Female	15,497,994	1	0.06		
fucking	Male	31,586,324	709	22.45	128.474	<0.001
	Female	15,497,994	132	8.52		
fucker(s)	Male	31,586,324	35	1.11	7.142	0.012
	Female	15,497,994	6	0.39		
All forms	Male	31,586,324	1322	41.85	162.124	<0.001
	Female	15,497,994	306	19.74		

As can be seen from Table 17, male authors use all forms of *fuck* more than twice as frequently as female authors. This difference is significant at the level p <0.001 (LL = 162.124, 1 d.f.). The difference between the two genders is also quantitatively significant for each word form, though the significance level may vary, with *fucking* demonstrating the greatest contrast. In terms of word forms, while female authors appear to prefer *fuck* to *fucking* more than male authors (see Table 19), the difference is not statistically significant (LL = 0.439, 1 d.f.). The proportion and rank of word forms show a very similar distribution pattern across author gender (Table 18). The fluctuation in the normalized frequencies can be discarded as they are not significant (LL = 1.162, 3 d.f.).

Table 18 Proportion and rank of word forms by male and female authors

Gender	*Form*	*Proportion (%)*	*Rank*
Male	*fucking*	53.63	1
	fuck	36.76	2
	fucked	5.90	3
	fucker(s)	2.65	4
	fucks	1.06	5
Female	*fucking*	43.14	2
	fuck	48.04	1
	fucked	6.54	3
	fucker(s)	1.96	4
	fucks	0.33	5

Table 19 Comparison of the normalized frequencies of word forms across gender

Form	*Male*	*Female*	*LL score*	*Sig. level*	*LL score*	*Sig. level*
fucking	22.45	8.52	0.439	0.570		
fuck	15.39	9.49				
fucked	2.47	1.29			1.162	0.867
fucker(s)	1.11	0.39	0.680	1.000		
fucks	0.44	0.06				

4.3 Age of author

Author age in writing is a sociolinguistic variable comparable to speaker age in speech and may, therefore, influence the distribution of *fuck*. Table 20 compares age groups of authors in the BNC written section by word form. As can be seen, the differences in the frequencies of *fuck* between authors of different age groups are statistically significant when all word forms are taken as a whole. An analysis by word form shows that except for the two very infrequent words *fucks* (3 instances) and *fucker(s)* (11 instances), all of the other word forms demonstrate a significant variation between age groups.

While young people also use *fuck* a lot in writing as they do in speech, the pattern of using *fuck* in writing appears to be different from that in speech in spite of some similarities, as shown in Table 21. In written English, the age group 60+ uses *fuck* least frequently. However, authors aged 25–34 use *fuck* most frequently, followed by the age group 45–59. While authors aged 45–59 use *fuck* slightly more often than those aged 35–44, the difference is not statistically significant (LL = 1.721, $p = 0.217$). Like speakers under 15, authors of the same age group use *fuck* more frequently than expected, though not as extensively as in speech. Surprisingly, people aged 15–24 use *fuck* less frequently than expected in written English, though this age group is the most frequent user of *fuck* in spoken English.

4.4 Gender of audience

The BNC classifies the gender of the intended audience of writing contained in the corpus into four types: male, female, mixed and unknown. In this article we consider only the first three categories. Table 22 compares the use of different word forms across gender.

As can be seen from Table 22, when all word forms are considered together, the difference between audience genders is statistically significant. However, *fucked* is the only word form which, in itself, shows a significant difference of distribution across writing intended for males and writing intended for females. *Fucked* is frequently used as the past form of the word with its literal meaning. Writing with an intended female audience contains significantly fewer occurrences of *fucked* than writings for an intended male audience. Other word

Table 20 Age of author

Form	*Age*	*Words*	*RF*	*NF*	*LL score*	*Sig.level*
	0–14	581,962	3	5.15		
	15–24	437,149	3	6.86		
fuck	25–34	1,325,516	97	73.18	178.234	<0.001
	35–44	2,813,226	32	11.37		
	45–59	2,847,335	36	12.64		
	60+	2,451,519	14	5.71		
	0–14	581,962	0	0		
	15–24	437,149	0	0		
fucked	25–34	1,325,516	20	15.09	46.263	<0.001
	35–44	2,813,226	5	1.78		
	45–59	2,847,335	11	3.86		
	60+	2,451,519	0	0		
	0–14	581,962	0	0		
	15–24	437,149	0	0		
fucks	25–34	1,325,516	1	0.75	3.286	0.778
	35–44	2,813,226	1	0.36		
	45–59	2,847,335	1	0.35		
	60+	2,451,519	0	0		
	0–14	581,962	12	20.62		
	15–24	437,149	5	11.44		
fucking	25–34	1,325,516	87	65.63	121.236	<0.001
	35–44	2,813,226	36	12.8		
	45–59	2,847,335	41	14.4		
	60+	2,451,519	21	8.57		
	0–14	581,962	2	3.44		
	15–24	437,149	0	0		
fucker(s)	25–34	1,325,516	3	2.66	7.216	0.129
	35–44	2,813,226	1	0.36		
	45–59	2,847,335	4	1.4		
	60+	2,451,519	1	0.41		
	0–14	581,962	17	29.21		
	15–24	437,149	8	18.3		
All forms	25–34	1,325,516	208	156.92	336.394	<0.001
	35–44	2,813,226	75	26.66		
	45–59	2,847,335	93	32.66		
	60+	2,451,519	36	14.68		

forms (especially *fuck* and *fucking*) used for emphasis (cf. section 5.1) do not show a significant contrast.

Interestingly, writing intended for a mixed audience is quite similar to writing intended for a male audience in terms of distribution patterns of *fuck* (LL = 0.134, d.f. = 1, $p = 0.714$) when all word forms are taken together. The difference in

Table 21 Comparison of speech and writing

	Spoken		*Written*	
Age group	*NF*	*Rank*	*NF*	*Rank*
0–14	851.01	2	29.21	3
15–24	1549.26	1	18.3	5
25–34	618.65	3	156.92	1
35–44	74.99	5	26.66	4
45–59	138.86	4	32.66	2
60+	18.71	6	14.68	6

Table 22 Gender of audience

Form	*Gender*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	Male	2,451,934	21	8.56	0.521	<0.471
fuck	Female	6,235,502	44	7.06		
	Mixed	54,289,029	591	10.89	–	–
	Male	2,451,934	17	6.93	28.091	<0.001
fucked	Female	6,235,502	3	0.48		
	Mixed	54,289,029	90	1.66	–	–
	Male	2,451,934	0	0	–	–
fucks	Female	6,235,502	0	0		
	Mixed	54,289,029	14	0.26	–	–
	Male	2,451,934	24	9.79	1.405	0.236
fucking	Female	6,235,502	45	7.22		
	Mixed	54,289,029	701	12.91	–	–
	Male	2,451,934	0	0	–	–
fucker(s)	Female	6,235,502	0	0		
	Mixed	54,289,029	43	0.79	–	–
	Male	2,451,934	62	25.29	10.270	0.001
All forms	Female	6,235,502	92	14.75		
	Mixed	54,289,029	1439	26.51	–	–

distributions of *fuck* in writing intended for females and that for a mixed audience is statistically significant at the level $p < 0.001$ (LL = 35.363, 1 d.f.). With respect to individual word forms, the difference between writing with an intended male audience and writing intended for a mixed audience is not statistically significant, while the difference between writing with an intended female audience and writing intended for a mixed audience is significant for *fuck* and *fucking*. For *fucked*, the difference of writing for the three types of audience is significant, though writing intended for a mixed audience is more akin to writing with an intended female audience.

4.5 Age of audience

This section examines the possible influence of audience age on the pattern of uses of *fuck* in written English. There are four age groups for audience: adults, teenagers, children and unknown. We consider the first three categories. Table 23 gives the frequencies of *fuck* across these age groups.

As can be seen from the table, writing for adults contains nearly twice as many uses of *fuck* as writing for teenagers. *Fuck* occurs in writing for adults more than seven times as frequently as in writing for children. This difference is significant at the level $p < 0.001$. In terms of word forms, the greatest contrast is for *fucking*, followed by *fuck*, while *fucked*, *fucks* and *fucker(s)* do not show a significant contrast because of the low overall frequencies of these word forms.[13] This finding is in line with the social convention that writing for children should avoid swear-words in order to discourage the use of this form of language by teenagers.[14]

Table 23 Age of audience

Form	*Age*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	Adult	82,335,639	784	9.52		
fuck	Teenager	1,697,721	10	5.89	14.482	0.001
	Child	969,382	1	1.03		
	Adult	82,335,639	128	1.55		
fucked	Teenager	1,697,721	2	1.18	0.755	0.712
	Child	969,382	0	0		
	Adult	82,335,639	18	0.22		
fucks	Teenager	1,697,721	0	0	0.110	1.000
	Child	969,382	0	0		
	Adult	82,335,639	960	11.66		
fucking	Teenager	1,697,721	7	4.12	22.217	<0.001
	Child	969,382	2	2.06		
	Adult	82,335,639	48	0.58		
fucker(s)	Teenager	1,697,721	2	1.18	1.412	0.347
	Child	969,382	0	0		
	Adult	82,335,639	1938	23.54		
All forms	Teenager	1,697,721	21	12.37	37.603	<0.001
	Child	969,382	3	3.09		

4.6 Level of audience

The BNC annotation scheme includes information pertaining to the levels of intended readership, thus enabling us to explore the pattern of uses of *fuck* along this dimension. Table 24 compares the distribution of *fuck* in writings for different levels of audience.

It can be seen that the rate of usage of *fuck* declines with a higher audience level. As far as word forms are concerned, the difference between audience levels

Table 24 Level of audience

Form	*Level*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
	Low	17,126,603	229	13.37	7.998	0.005
fuck	Medium	43,837,214	465	10.61	118.407	<0.001
	High	23,967,568	101	4.21	–	–
	Low	17,126,603	32	1.87	0.086	0.660
fucked	Medium	43,837,214	77	1.76	10.527	0.005
	High	23,967,568	21	0.88	–	–
	Low	17,126,603	5	0.29	0.384	0.826
fucks	Medium	43,837,214	9	0.21	0.853	0.671
	High	23,967,568	4	0.17	–	–
	Low	17,126,603	243	14.19	2.73	0.098
fucking	Medium	43,837,214	547	12.48	52.212	<0.001
	High	23,967,568	179	7.47	–	–
	Low	17,126,603	13	0.76	0.001	0.980
fucker(s)	Medium	43,837,214	33	0.75	12.749	0.002
	High	23,967,568	4	0.17	–	–
	Low	17,126,603	522	30.48	9.711	0.002
All forms	Medium	43,837,214	1131	25.8	178.857	<0.001
	High	23,967,568	309	12.89	–	–

is statistically significant for all word forms except *fucks*, which occurs only rarely. The greatest contrast is found for *fuck* (LL = 118.407). It is also interesting to note that medium level is closer to low level than it is to high level. Except for *fuck*, the difference between different audience levels is not quantitatively significant. While it is not clear why the word form *fuck* shows a significant contrast, we speculate that this is due to its high overall frequency. When all word forms are taken as a whole, the difference between medium and low levels is significant (LL = 9.711, 1 d.f.). But this significance is probably skewed by the marked contrast for the word form *fuck*.

4.7 Reception status

In this section we examine the potential relationship between reception status and the pattern of usage of *fuck*. The BNC classifies written texts into four types in terms of their reception status: high, medium, low and unknown. We discard cases where reception status is unknown. As can be seen from Table 25, whether we consider the word forms of *fuck* separately or together, the difference in the distribution of *fuck* across reception status is statistically significant. In this case, medium reception status appears to be closer to high than low status. In terms of word forms, the difference between high and medium reception statuses is significant only for *fucks* and *fucking*.

Table 25 Reception status

Form	*Level*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*	*LL score*	*Sig. level*
	High	24,138,350	278	11.52	1.353	0.245		
fuck	Medium	31,885,282	402	12.61			73.179	<0.001
	Low	16,488,041	83	5.03	–	–		
	High	24,138,350	40	1.66	0.776	0.381		
fucked	Medium	31,885,282	63	1.98			8.456	0.015
	Low	16,488,041	15	0.91	–	–		
	High	24,138,350	11	0.46	7.357	0.007		
fucks	Medium	31,885,282	3	0.09			7.077	0.025
	Low	16,488,041	4	0.24	–	–		
	High	24,138,350	402	16.65	6.252	0.012		
fucking	Medium	31,885,282	447	14.02			179.914	<0.001
	Low	16,488,041	60	3.64	–	–		
	High	24,138,350	13	0.54	3.006	0.083		
fucker(s)	Medium	31,885,282	30	0.94			9.681	0.008
	Low	16,488,041	4	0.24	–	–		
	High	24,138,350	744	30.82	0.639	0.424		
All forms	Medium	31,885,282	945	29.64			245.785	<0.001
	Low	16,488,041	166	10.07	–	–		

Table 26 Distribution pattern of *fuck* by reception status

Row	*Form*	*High*	*Medium*	*Low*
1	*fuck*	2	1	3
2	*fucked*	2	1	3
3	*fucks*	1	3	2
4	*fucking*	1	2	3
5	*fucker(s)*	2	1	3
6	All forms	1	2	3

We can get a vague picture of the pattern of usage of *fuck* across reception status by sorting by normalized frequencies, as shown in Table 26. The table by itself does not show a pattern of *fuck* usage. However, if we combine Tables 25 and 26 and take statistical significance into consideration, we are able to see clearly the pattern of usage for *fuck* across reception status.

Table 25 shows that the difference between high and medium reception statuses is not statistically significant for *fuck* ($p = 0.245$), *fucked* ($p = 0.381$) and *fucker* ($p = 0.083$), hence the *High* and *Medium* in rows 1, 2 and 5 in Table 26 could be swapped, to retain the order *High (1)*, *Medium (2)* and *Low (3)*. Note, however, that the ranks of *High* and *Medium* cannot be inverted on rows 3 and 4, as the inverted order would run counter to the significance tests for these words in

Table 25 which show that these words are used more in *High* contexts. However, in row 3 *Medium* and *Low* could be swapped to become *High (1)*, *Medium (2)* and *Low (3)* because the difference between these two categories is not statistically significant (LL = 1.551, 1 d.f., $p = 0.213$). If all the possible rearrangements discussed are undertaken, the pattern of usage of *fuck* across reception status becomes *High>Medium>Low* for each word form and all word forms combined. This finding is unusual and the explanation for this phenomenon is beyond the corpus-based approach; it would, in our opinion, require substantial sociological study to explain.

Table 27 Medium of text

Form	*Medium*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*	*LL score*	*Sig. level*
	Book	52,574,506	667	12.69	0.198	0.657		
	Misc. unpub.	3,461,953	47	13.58				
fuck	Periodical	23,978,695	80	3.34	–	–	265.830	<0.001
	Misc. pub.	3,922,977	1	0.25	–	–		
	To-be-spoken	861,592	0	0	–	–		
	Book	52,574,506	100	1.9	0.740	0.390		
	Misc. unpub.	3,461,953	9	2.6				
fucked	Periodical	23,978,695	19	0.79	–	–	22.373	<0.001
	Misc. pub.	3,922,977	1	0.25	–	–		
	To-be-spoken	861,592	0	0	–	–		
	Book	52,574,506	16	0.3	0.619	1.000		
	Misc. unpub.	3,461,953	2	0.58				
fucks	Periodical	23,978,695	0	0	–	–	11.720	0.014
	Misc. pub.	3,922,977	0	0	–	–		
	To-be-spoken	861,592	0	0	–	–		
	Book	52,574,506	875	16.64	22.333	<0.001		
	Misc. unpub.	3,461,953	25	7.22				
fucking	Periodical	23,978,695	59	2.46	–	–	430/306	<0.001
	Misc. pub.	3,922,977	8	2.04	–	–		
	To-be-spoken	861,592	0	0	–	–		
	Book	52,574,506	41	0.78	0.030	1.000		
	Misc. unpub.	3,461,953	3	0.87				
fucker(s)	Periodical	23,978,695	6	0.25	–	–	11.007	0.018
	Misc. pub.	3,922,977	0	0	–	–		
	To-be-spoken	861,592	0	0	–	–		
	Book	52,574,506	1699	32.32	6.137	0.013		
	Misc. unpub.	3,461,953	86	24.84				
All forms	Periodical	23,978,695	164	6.84	–	–	709.749	<0.001
	Misc. pub.	3,922,977	10	2.55	–	–		
	To-be-spoken	861,592	0	0	–	–		

4.8 Medium of text

Five basic types of medium of text are annotated in the BNC corpus, *book*, *miscellaneous unpublished*, *periodical*, *miscellaneous published* and *written-to-be-spoken*. This section uses this information to examine the effect of publication medium on the distribution pattern of *fuck*. Table 27 compares the rate of usage of *fuck* across medium. It is clear that the contrast between types of medium is statistically significant for all the word forms. While *miscellaneous unpublished* ranks before *book* for four out of five word forms (*fuck*, *fucked*, *fucks* and *fucker[s]*), the difference in the frequencies between the two media is not statistically significant. Hence, for these word forms *book* and *miscellaneous unpublished* could be re-ordered. *Book* ranks before *miscellaneous unpublished* only for *fucking*. Yet the difference in its frequency between the two types of medium is significant, therefore *book* and *miscellaneous unpublished* cannot be re-ordered. Word forms of *fuck* are most frequently used in *book*, followed by *miscellaneous unpublished*, *periodical*, *miscellaneous published* and *written-to-be-spoken*. As can be seen from the table, forms of *fuck* occur nearly five times as frequently in books as in periodicals, and over 12 times as frequently as in miscellaneous published works. No use of *fuck* is found in written-to-be-spoken scripts.

4.9 Date of creation

In this section we compare written English in the periods 1960–74 and 1975–93 to see whether the pattern of uses of *fuck* in written British English has changed. As date of creation is encoded for the written section of the BNC alone, it is not possible to examine changes in the distribution pattern of *fuck* in spoken English using the BNC. As there is no ready-made analogue of the spoken BNC available for an earlier period, the exploration of diachronic change in spoken English is, in effect, impossible using the corpus-based methodology.

Table 28 Date of creation

Form	*Date*	*Words*	*RF*	*NF*	*LL score*	*Sig. level*
fuck	1975–93	75,501,632	762	10.09	5.241	0.022
	1960–74	2,036,939	11	5.4		
fucked	1975–93	75,501,632	128	1.7	6.815	0.009
	1960–74	2,036,939	0	0		
fucks	1975–93	75,501,632	18	0.24	0.958	1.000
	1960–74	2,036,939	0	0		
fucking	1975–93	75,501,632	937	12.41	0.020	0.888
	1960–74	2,036,939	26	12.76		
fucker(s)	1975–93	75,501,632	47	0.62	1.642	0.200
	1960–74	2,036,939	3	1.47		
All forms	1975–93	75,501,632	1892	25.06	2.520	0.112
	1960–74	2,036,939	40	19.64		

As can be seen in Table 28, when all word forms are taken together there is no significant difference in the frequency of *fuck* in the two periods under consideration, in spite of a 5 percent increase in usage in the period 1975–93.[15] In terms of word forms, however, there are some remarkable changes. While *fucking* was used at almost exactly the same rate in the two periods, the frequency of the form *fuck* doubled in the later period. The difference in the frequencies of *fucker(s)* is not significant, but the use of the word was reduced by half in 1975–93. It is also interesting to note that the use of the words *fucked* and *fucks* appears to be a new development in written English in the period 1975–93, because the texts sampled for 1960–75, amounting to 2,000,000 words, do not contain a single instance of the two words. For the moment we simply note this phenomenon, though we will return to consider it in section 5.8.

5 Categorization of *fuck*

In this section we apply the category scheme developed for the Lancaster Corpus of Abuse (LCA) in McEnery et al. (2000: 45) to *fuck* in the whole BNC corpus. Our implementation of the scheme, however, is slightly different. The original annotation scheme consisted of 16 categories, but some of them (e.g. A, M, R and T) do not apply to *fuck*.[16] In our revised annotation scheme, the B category (adverbial booster, as in *Fucking marvellous*) is folded into E (emphatic intensifier, see Table 29). We made this decision because both categories provide emphasis. The only difference between the B and E categories is the part of speech of the word following *fucking*, yet in many cases (particularly in the structure *fucking* + *adjective* + *noun*) it is difficult for a human analyst to make the distinction. We also allowed the original N category (premodifying negative adjective, as in *the fucking idiot*) to be folded into E for ease of annotation, as the semantic distinction focused on the modified head noun can be difficult to make. Consider *it's only a fucking Saturday job*. It is a matter of debate whether *fucking*

Table 29 Category scheme for swear words

Code	*Description*	*Examples*
G	General expletive	(Oh) fuck!
P	Personal insult referring to defined entity	You fuck!/that fuck
C	Cursing expletive	Fuck you!/me!/him!/it!
D	Destinational usage	Fuck off!/he fucked off
L	Literal usage denoting taboo referent	He fucked her
E	Emphatic intensifier	Fucking marvellous!/in the fucking car
O	'Pronominal' form	Like fuck/fat as fuck
I	Idiomatic 'set phrase'	Fuck all/give a fuck/thank fuck
X	Metalinguistic or unclassifiable due to insufficient context	The use of the word "fuck"/you never fucking

here is an N or an E, and the categorization would be determined solely by the highly subjective attitude of the annotator to *Saturday job*. It was because such unclear cases were far from rare when we annotated the corpus data that we abandoned the N category. We joined the F category (figurative extension of literal use, as in *to fuck about*) with the I category because all the expressions such as *fuck about*/*around*/*up* can reasonably be considered to be idiomatic usage. Our revised version of the annotation scheme has nine categories. These annotations were applied manually to all word forms of *fuck* in the BNC.[17] Table 29 lists these categories of *fuck* and gives typical examples from the BNC.

Table 30 *Fuck* and its morphological variants in the BNC corpus

Category	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*X*	*Total*
Frequency	323	90	288	261	344	74	2684	591	151	4806
Percent	6.72	1.87	5.99	5.43	7.16	1.54	55.85	12.30	3.14	100

Table 30 shows the frequencies and proportions of *fuck* and its morphological variants in the BNC. As can be seen in the table, these are most frequently used as an emphatic intensifier (category E), followed by idiomatic use (category I). In the sections that follow, we explore the categories of *fuck* by using the major parameters encoded in the BNC. Note, however, that we use only the first eight categories, discarding the 151 instances of the X category.[18]

5.1 Variation across word form

As noted in sections 3.4 and 4.4, *fucked* and *fucks* are very frequently used to denote a literal meaning whereas *fucking* is most frequently used for emphasis. Table 31 shows the distributions of different word forms across category of usage.

It can be seen that the difference in the distributions is statistically significant at the level $p < 0.001$. *Fuck* is most frequently used idiomatically, as in *what the*

Table 31 Word form vs usage category

Form	*Total*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*LL score*	*Sig. level*
fuck	1362	227	15	278	247	149	31	5	410		
fucked	192	0	0	0	6	69	0	0	117		
fucks	28	0	1	0	2	16	0	0	9		
fucking	2998	96	3	10	6	108	43	2677	55	5320.1	<0.001
fucker(s)	75	0	71	0	0	2	0	2	0		
Total	4655	323	90	288	261	344	74	2684	591		
Percent	100	6.94	1.93	6.19	5.61	7.39	1.59	57.66	12.69		

fuck, *for fuck's sake*, *give a fuck* and *fuck up/around/about*. It is interesting to note that *the fuck* often goes with a *wh*-word like *what*, *who*, *where*, *how* and *why* (189 out of 230 instances of *the fuck* of category I) while *give a fuck* generally occurs with a negative particle or pronoun like *not* or *nobody* (18 out of 24 instances) or in a question (6 out of 24 instances). The word form *fuck* is also used very frequently as a cursing (e.g. *Fuck you!*), destinational (e.g. *Fuck off!*), or general expletive (e.g. *Oh fuck!*), but is used infrequently (only 5 instances) as an emphatic intensifier (e.g. *I am not fuck laughing, am I?*).[19] In contrast, *fucking* is used most frequently for emphasis but least frequently as a personal insult or destinational expletive. Interestingly, the emphatic *fucking* is sometimes used (20 out of 2677 instances of *fucking* of the E category) as an infix, splitting a whole word (e.g. *every fucking where*) or two parts of a name (e.g. *Doctor fucking Halziel* and *Jesus fucking Christ*). The most frequent use of *fucked* is idiomatic, mainly in the structure *be/get fucked (up)*, though the meaning of *fucked* is often referential (e.g. *She fucked him with enthusiasm*). The frequency of *fucks* is low. In addition to its literal usage (e.g. *the king who fucks his daughter*), *fucks* is often used idiomatically, as in *Nobody fucks with us* and *Cos it fucks it all up*. The most important use of *fucker(s)* is as a personal insult (e.g. *You stupid fucker* and *You let them fuckers in?*). Note that *fuck(s)* has begun to be used sometimes to replace *fucker(s)*, e.g. *you little fuck* and *Those fucks don't make cars, they make toys*.

5.2 Variation across register

We noted in section 2 that *fuck* and its variants occur significantly more frequently in spoken English than in written English. In addition to this quantitative difference, spoken and written registers also differ qualitatively in terms of the use of *fuck*. As can be seen in Table 32, while category E is the most

Table 32 Difference in the uses of *fuck* across register

Form	*Register*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
fuck	Spoken	162	4	104	120	26	14	5	139	574
	Written	65	11	174	127	123	17	0	271	788
fucked	Spoken	0	0	0	4	11	0	0	47	62
	Written	0	0	0	2	58	0	0	70	130
fucks	Spoken	0	0	0	1	5	0	0	4	10
	Written	0	1	0	1	11	0	0	5	18
fucking	Spoken	94	3	7	5	41	42	1810	30	2032
	Written	2	0	3	1	67	1	867	25	966
fucker(s)	Spoken	0	23	0	0	0	0	2	0	25
	Written	0	48	0	0	2	0	0	0	50
Total (spoken)		256	30	111	130	83	56	1817	220	2703
Rank/% (spoken)		2/9.47	8/1.11	5/4.11	4/4.81	6/3.07	7/2.07	1/67.22	3/8.14	100
Total (written)		67	60	177	131	261	18	867	371	1952
Rank/% (written)		6/3.43	7/3.07	4/9.07	5/6.71	3/13.37	8/0.92	1/44.42	2/19.01	100

frequent category in both spoken and written registers, the frequency of this category is considerably higher in speech than in writing. Category I is very frequent in both written and spoken registers, but the number in this category is much higher in writing than in speech, suggesting that the former is more formal and elaborate than the latter. In contrast, category G is more frequent in speech, which is in harmony with its informal style. The L category ranks third, following categories E and I, in the written register whereas it ranks fifth in the spoken register. The frequency of this category is much higher in the written register than in the spoken register. This suggests that when people say *fuck* in speech, it is most likely that they want to show their anger or annoyance. When people use *fuck* in writing, they more often refer to coitus. It is also clear that in both speech and writing *fucker(s)* is mainly used as a personal insult (e.g. *Oh you fucker!*).[20]

5.3 User gender

Sections 3.4 and 4.2 show that in both spoken and written registers, while males use *fuck* much more frequently than females, the distribution pattern of word forms across gender is quite similar. In this section we compare the distributions of usage categories across gender in both spoken and written registers.

Table 33 Difference in the uses of *fuck* across speaker gender

Form	*Gender*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
fuck	Male	102	2	59	66	6	11	1	88	335
	Female	21	0	27	20	4	1	3	26	102
fucked	Male	0	0	0	2	5	0	0	18	25
	Female	0	0	0	1	1	0	0	11	13
fucks	Male	0	0	0	0	4	0	0	1	5
	Female	0	0	0	0	1	0	0	1	2
fucking	Male	64	2	6	4	30	32	1156	23	1317
	Female	8	1	0	0	4	3	287	1	304
fucker(s)	Male	0	18	0	0	0	0	0	0	18
	Female	0	0	0	0	0	0	2	0	2
Total (male)		166	22	65	72	45	43	1157	130	1700
Rank/% (male)		2/9.76	8/1.29	5/3.82	4/4.24	6/2.65	7/2.53	1/68.06	3/7.65	100
Total (female)		29	1	27	21	10	4	292	39	423
Rank/% (female)		3/6.86	8/0.24	4/6.38	5/4.96	6/2.36	7/0.95	1/69.03	2/9.22	100

As can be seen in Table 33, categories of *fuck* distribute, in speech; in a similar pattern for male and female speakers in terms of both rank and proportions, though males appear to use categories G, P and O more frequently whereas females use the C and I categories more frequently.

Table 34 shows that in writing, for both male and female authors, the most frequent use is also the E category, followed by categories I and L. However, male

Table 34 Difference in the uses of *fuck* across author gender

Form	*Gender*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
fuck	Male	36	11	95	71	75	14	2	179	483
	Female	18	0	29	24	32	2	0	41	146
fucked	Male	0	0	0	1	40	0	0	37	78
	Female	0	0	0	0	11	0	0	9	20
fucks	Male	0	1	0	0	10	0	0	3	14
	Female	0	0	0	0	1	0	0	0	1
fucking	Male	1	0	2	1	37	0	650	15	706
	Female	0	0	0	0	15	0	110	7	132
fucker(s)	Male	0	34	0	0	1	0	0	0	35
	Female	0	6	0	0	0	0	0	0	6
Total (male)		37	46	97	73	163	14	652	234	1316
Rank/% (male)		7/2.81	6/3.50	4/7.37	5/5.55	3/12.39	8/1.07	1/49.54	2/17.78	100
Total (female)		18	6	29	24	59	2	110	57	305
Rank/% (female)		6/5.90	7/1.97	4/9.51	5/7.87	2/19.34	8/0.66	1/36.06	3/18.69	100

authors use *fuck* more frequently than female authors for emphasis while female authors use *fuck* more frequently to refer to coitus or as a general expletive.

A comparison of the rank and proportions of usage categories for male and female users in speech and writing tells us more about the gender differences across the two registers. As Tables 33 and 34 show, in spoken English there is almost no difference between the two genders for the E category. This is because using *fuck* for emphasis is the most important usage in the spoken register and both genders use this category very frequently (nearly 70 percent of all uses). In writing, however, the distribution of *fuck* across the categories is more balanced. In addition to providing emphasis, *fuck* is often used idiomatically in writing. The I category accounts for a much lower frequency in speech than in writing, whereas E and G show higher frequencies than in writing. While both female and male users are more likely to use *fuck* in its L form in writing, it is in speech that male users are more likely than female users to use *fuck* as a general expletive.

5.4 User age

This section examines the distribution of the eight categories of *fuck* across speaker age in speech and author age in writing. Table 35 shows the frequency, rank and proportion of the categories used by speakers of different age groups. As can be seen, except for age groups 35–44 and 60+, the eight categories of *fuck* distribute across age groups in a very similar way in terms of rank. Category E is the most frequent, followed by I and G. Category P is the most infrequent for all age groups except 35–44. Age groups 0–14 and 35–44 and 60+ appear to be atypical in their uses of *fuck*. For children under 15, the proportion of the E

Table 35 Speaker age

Age	*Data type*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
0–	Frequency	60	2	35	22	20	4	191	40	374
14	Rank/%	2/16.04	8/0.53	4/9.36	5/5.88	6/5.35	7/1.07	1/51.07	3/10.70	100
15–	Frequency	67	9	21	40	14	25	521	64	761
24	Rank/%	2/8.80	8/1.18	6/2.76	4/5.26	7/1.84	5/3.29	1/68.46	3/8.41	100
25–	Frequency	40	9	19	25	15	13	488	44	653
34	Rank/%	3/6.13	8/1.38	5/2.91	4/3.83	6/2.30	7/1.99	1/74.73	2/6.74	100
35–	Frequency	2	1	5	2	0	1	65	1	77
44	Rank/%	3=/2.60	5=/1.30	2=/6.49	3=/2.60	8/–	5=/1.30	1/84.42	5=/1.30	100
45–	Frequency	25	2	10	3	5	4	151	12	212
59	Rank/%	2/11.79	8/0.94	4/4.72	7/1.42	5/2.36	6/1.89	1/71.23	3/5.66	100
60+	Frequency	0	0	0	0	1	0	17	3	21
	Rank/%	4=/–	4=/–	4=/–	4=/–	3/4.76	4=/–	1/80.95	2/14.29	100

Table 36 Author age

Age	*Data type*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
0–14	Frequency	1	2	0	0	3	0	10	1	17
	Rank	4=	3	6=	6=	2	6=	1	4=	
	%	5.88	11.76	–	–	17.65	–	58.82	5.88	100
15–24	Frequency	0	0	0	0	1	0	4	3	8
	Rank	4=	4=	4=	4=	3	4=	1	2	
	%	–	–	–	–	12.50	–	50.00	37.50	100
25–34	Frequency	11	3	16	19	44	3	72	37	205
	Rank	6	7=	5	4	2	7=	1	3	
	%	5.37	1.46	7.80	9.27	21.46	1.46	35.12	18.05	100
35–44	Frequency	2	0	4	6	14	0	31	17	74
	Rank	6	7=	5	4	3	7=	1	2	
	%	2.70	–	5.41	8.11	18.92	–	41.89	22.97	100
45–59	Frequency	2	4	14	11	12	0	32	18	93
	Rank	7	6	3	5	4	8	1	2	
	%	2.15	4.30	15.05	11.83	12.90	–	34.41	19.35	100
60+	Frequency	0	1	6	4	3	0	20	2	36
	Rank	7=	6	2	3	4	7=	1	5	
	%	–	2.78	16.67	11.11	8.33	–	55.56	5.56	100

category is lower than in other age groups whereas the frequencies of categories G, C and L are much higher. Speakers over 60 appear to use *fuck* primarily for emphasis or idiomatically. Those aged 35–44 have a much higher proportion of the E category whereas they use categories G and I infrequently.

In writing, as in speech, the age groups 0–14 and 60+ are atypical in their uses of *fuck* (see Table 36). The two groups show a much lower frequency of the I category whereas they use category E more frequently than other age groups. The E category is the most frequent for all age groups, followed by I and L. In contrast

with speech, the L category is significantly more frequent in writing. This finding is in line with our conclusion in section 5.2.

5.5 Speaker social class

We noted in section 3.6 that speakers from the classes of C2 and DE are the most frequent users of *fuck*. As can be seen from Table 37, the two classes also use *fuck* similarly in terms of rank and proportion. While people from the classes AB and C1 also demonstrate some similarities (in categories P, C and O), class C1 uses the E category much more frequently whereas class AB uses the G, D and L categories more frequently. But for all social classes the E category is the most frequent category of *fuck*, followed by I and G.

Table 37 Speaker social class

Class	*Data type*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
AB	Frequency	36	0	14	15	14	7	163	38	287
	Rank/%	3/12.54	8/–	5=/4.88	4/5.23	5=/4.88	7/2.44	1/56.79	2/13.24	100
C1	Frequency	2	0	1	0	0	2	35	3	43
	Rank/%	3=/4.65	6=/–	5/2.33	6=/–	6=/–	3=/4.65	1/81.39	2/6.98	100
C2	Frequency	21	4	13	19	5	6	258	21	347
	Rank/%	2=/6.05	8/1.15	5/3.75	4/5.48	7/1.44	6/1.73	1/74.35	2=/6.05	100
DE	Frequency	25	4	12	5	5	6	174	20	251
	Rank/%	2/9.96	8/1.60	4/4.78	6=/1.99	6=/1.99	5/2.39	1/69.32	3/7.97	100

5.6 Speaker education level

In section 3.7, we noted that people who received less education say *fuck* more frequently. While it appears that people who received more education use the G and E categories less frequently and the I category more frequently, the influence of speaker education level on the uses of *fuck* is not very clear. This is perhaps due to relatively sparse data. As can be seen in Table 38, the frequencies of *fuck* for three out of the four groups are very low.

Table 38 Speaker education level

Edu.	*Data*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
Left sch.	Frequency	0	0	0	0	1	0	7	1	9
15–	Rank/%	4=/–	4=/–	4=/–	4=/–	2=/11.11	4=/–	1/77.78	2=/11.11	100
Left sch.	Frequency	46	6	22	21	9	11	423	38	576
15/16	Rank/%	2/7.99	8/1.04	4/3.82	5/3.64	7/1.56	6/1.91	1/73.44	3/6.60	100
Left sch.	Frequency	2	0	1	1	1	3	21	2	31
17/18	Rank/%	3=/6.45	8/–	5=/3.23	5=/3.23	5=/3.23	2/9.67	1/67.74	3=/6.45	100
Left sch.	Frequency	0	0	2	2	0	0	10	2	16
19+	Rank/%	5=/–	5=/–	2=/12.50	2=/12.50	5=/–	5=/–	1/62.50	2=/12.50	100

What we know for sure is that for people of all education levels the E and I categories are the most frequent uses of *fuck* whereas category P is the most infrequent. Interestingly, the distribution of these eight categories for those who left school before 15 years of age is very similar to the distribution pattern for speakers aged 60+ (see Table 35), because those who left school under 15 are mostly older people (cf. section 3.7).

5.7 Social contexts

In section 3.1, we noted that context-governed and demographically sampled speech differ significantly in their frequency of usage of *fuck*. This difference exists for two reasons: the first type of speech is more formal than the second (cf. Andersson and Trudgill, 1992: 60) and the first type occurs on public occasions whereas the second occurs on private occasions. These two factors also influence the distribution of uses of *fuck*.

As can be seen from Table 39, the most frequent uses of *fuck* in formal language used on public occasions are the E and I categories, whereas personal insult (P), curse (C), literal usage (L) and pronominal form (O) are non-existent. In informal language used on private occasions, in contrast, the distribution of *fuck* is more balanced across category. While the most important use is category E, the other categories like G and I are also used frequently.

Table 39 Social contexts

Form	*Context*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
fuck	Public	1	0	0	1	0	0	0	5	7
	Private	161	4	104	119	26	14	5	134	567
fucked	Public	0	0	0	0	0	0	0	1	1
	Private	0	0	0	4	11	0	0	46	61
fucks	Public	0	0	0	0	0	0	0	0	0
	Private	0	0	0	1	5	0	0	4	10
fucking	Public	0	0	0	1	0	0	12	2	15
	Private	94	3	7	4	41	42	1798	28	2017
fucker(s)	Public	0	0	0	0	0	0	0	0	0
	Private	0	23	0	0	0	0	2	0	25
Total (public)		1	0	0	2	0	0	12	8	23
Rank/% (public)		4/4.35	5=/–	5=/–	3/8.70	5=/–	5=/–	1/52.17	2/34.78	100
Total (private)		255	30	111	128	83	56	1805	212	2680
Rank/% (private)		2/9.51	8/1.12	5/4.14	4/4.78	6/3.10	7/2.09	1/67.35	3/7.91	100

5.8 Date of creation

We noted in section 4.9 that the frequencies of word forms *fuck* and *fucked* used in 1975–92 are significantly higher than in 1960–74. This section compares the

Table 40 Date of creation

Form	*Date*	*G*	*P*	*C*	*D*	*L*	*O*	*E*	*I*	*Total*
fuck	1960–74	1	0	3	2	1	0	0	1	8
	1975–92	61	11	163	121	119	16	0	267	758
fucked	1960–74	0	0	0	0	0	0	0	0	0
	1975–92	0	0	0	2	57	0	0	69	128
fucks	1960–74	0	0	0	0	0	0	0	0	0
	1975–92	0	1	0	1	11	0	0	5	18
fucking	1960–74	0	0	0	0	2	0	22	2	26
	1975–92	2	0	3	1	64	1	840	23	934
fucker(s)	1960–74	0	3	0	0	0	0	0	0	3
	1975–92	0	45	0	0	2	0	0	0	47
Total (1960–74)		1	3	3	2	3	0	22	3	37
Total (1975–92)		63	57	166	125	253	17	840	364	1885

Table 41 Significant difference between 1960–74 and 1975–92

Form	*Category*	*Date*	*Word number*	*Frequency*	*LL score*	*Sig. level*
fuck	I	1960–74	2,036,939	1	8.32	0.004
		1975–92	75,501,632	267		
fucked	I	1960–74	2,036,939	0	3.67	0.055
		1975–92	75,501,632	69		
All forms	I	1960–74	2,036,939	3	6.40	0.011
		1975–92	75,501,632	364		

uses of *fuck* in the two sampling periods. Table 40 gives the frequencies of usage categories.

On the basis of log-likelihood tests using word numbers for data of the two sampling periods and the frequencies in Table 40, we found that, except for the I category, the difference in the two periods is not statistically significant. Table 41 shows all the cases where a difference is highly or marginally significant. As can be seen, with respect to individual word forms only the difference for the I category of *fuck* and *fucked* between the two sampling periods is significant. When all word forms are taken together, the difference between the two periods is again significant only for the I category.

A careful examination of the uses of the I category in the two periods shows a potential qualitative difference. As can be seen in Table 42, only 3 forms of category I were used in 1960–74 whereas in 1975–92 10 forms were used, most of which were unseen in the earlier period. One must conclude that either there has been a proliferation in the usage and numbers of I forms, or these forms were previously the object of much more censorship in their written form than they currently are.

Table 42 Uses of category I in 1960–74 and 1975–92

No.	*Forms*	*1960–74*	*1975–92*
1	*Wh-word the fuck*	1	126
2	*the fuck*	0	25
3	the *fucking God*/*God fucking* group (*for ~'s sake, ~ knows, thank ~, thank the ~ing, in ~'s name, in the ~*)	0	38
4	*give*/*care a fuck*	0	17
5	*fuck all*	0	11
6	*fuck (up*/*off)*	0	69
7	*fuck with*	0	12
8	*fuck about*/*around*/*out*	1	12
9	*be*/*get fucked (up*/*out*/*off)*	0	45
10	*fucking well*	1	9
	Total	3	364

6 Conclusion – insights into corpus linguistics

In this article we have used the information encoded in the BNC metadata to explore the distribution pattern of *fuck* and its variants both within and across speech and writing. We have also applied a categorization scheme to 4806 examples of *fuck* and its associated word forms in the BNC. Hence, while this article focuses on a systematic account of the use of *fuck* in modern British English from a sociolinguistic point of view, it also provides a usage-based account of *fuck* by classifying it into nine categories. While this article studies only *fuck*, the methodology used in this work could also be applied to the investigation of other swear-words, and should provide a fruitful avenue of future research.

While the investigation presented in this article is only possible with appropriate corpus resources, we feel that corpus linguistics, as a methodology, cannot answer all questions of interest to linguists studying swearing (cf. McEnery et al., 2000: 47). The BNC has allowed us an unprecedented insight into how the word *fuck* is used in English. That insight could be extended to other swear-words. Yet the study has also shown that the use of the corpus as a methodology defines the boundaries of any given study. The corpus can limit and define the range of explanations which we may be able to propose for any given observation, but it cannot, in itself, select and propose a unique explanation with ease, if at all. On the other hand it provides an excellent descriptive methodology. But moving beyond description to explanation brings us back to intuition. Human intuition has a role to play in corpus linguistics. The corpus allows us to discount hypotheses, and at times can provide results which run counter to our intuition (for example, see the results in section 4.7). But corpora do not provide explanations for what we see. These explanations must be developed using other methodologies. Hence this article has helped to refine the view of what a corpus can and cannot do. A corpus can assist greatly in the testing of hypotheses and the establishment of the parameters within which any explanation of some feature of

language use must work. As such it is a useful methodology in linguistics. Yet it cannot explain the findings it yields. This remains the task of the human analyst, drawing upon a wide range of resources – and methodologies.

Notes

1 Though McEnery (2004) seeks to provide a better historical explanation of attitudes to bad language in English.
2 McEnery et al. (2000) is an example along this line, in which the authors provide an account of the distribution pattern of *fuck* across speaker gender, age and social class in the spoken section of the BNC corpus.
3 The BNC here refers to the web-based British National Corpus Version 2, from which all of the examples in this article are cited.
4 Unless otherwise stated, we use *fuck* to refer to the word form *fuck* and all its morphological variants including *fuck, fucked, fucks, fuckin(g)* and *fucker(s)*. There are 4806 such instances in the BNC. Note also that when we refer to the word form *fucking* hereafter in this article, we mean either *fuckin'* or *fucking*.
5 In the tables in this article, *RF* refers to raw frequency while *NF* refers to normalized frequency, i.e. frequency per 1,000,000 words.
6 The log-likelihood score is used in this article to determine the significance of observed pairwise differences. The critical value for statistical significance at $p < 0.05$ is 3.84. The critical value for statistical significance at $p < 0.001$ is 10.83.
7 While Stenström's (1991) observation may be true of the LLC corpus, it does not necessarily apply to the spoken section of the BNC. The normalized frequencies (per million words) we obtained for *god*/*heaven* and *damn*/*hell* from the spoken BNC are as follows: *god*/*heaven* (female: 489.94, male: 376.16), *damn*/*hell* (female: 178.77, male: 145.99). It is clear that female speakers use both groups of words significantly more frequently than male speakers. For the *heaven* group, the log-likelihood score is 59.227 (1 d.f.), and for the *hell* group 13.073 (1 d.f.), both significant at the level of $p < 0.001$. A plausible explanation for the difference between Stenström's observation and our finding is diachronic change – the spoken material in the LLC was primarily collected in the 1960s and 1970s whereas the spoken section of the BNC contains spoken material that was mainly collected in the early 1990s. It may be that, as Rosalind Coward suggests, 'women are now talking seriously dirty' (*New Statesman & Society*, 9 June 1989: 42, cited in Hughes, 1991: 211) and that this is evidence for an increase of swearing amongst women.
8 In this article we use the metadata encoded in the BNC for *speaker* rather than *respondent*, as a respondent was a recruit who used a portable tape recorder to record their own speech and the speech of people they conversed with, while a speaker was the person who actually produced a particular utterance (cf. Aston and Burnard, 1998: 32).
9 We doubt a statistical test based on such a low frequency (only five instances of *fucks*) can yield a reliable result.
10 A popular explanation for swearing is that people use swear-words when they do not have other words at their disposal, i.e. their vocabulary is so small that they have to use 'easy' and 'lazy' words in certain situations (cf. Andersson and Trudgill, 1992: 65).
11 As well as the four categories shown, the BNC also uses the category *Still in education*. We decided to exclude this category from our analysis because this group contains members who will eventually integrate into the other groups. There are 807.74 instances of *fuck* per million words (443 instances in 548,444 words) for those still in education. Interestingly, people of this group use *fuck* frequently as they are mostly of the age group 15–24.
12 Note that the reason for this is that the school-leaving age was raised to 16 in the UK in 1973 by Margaret Thatcher, then Education Secretary. Only older speakers could have left school younger than 16.
13 There are only 2.73, 0.22 and 1.76 instances of *fucked*, *fucks* and *fucker(s)* per million words.
14 This is also, in Britain, the rationale for the so-called watershed time of 9 p.m., before which bad language is not broadcast, a convention adhered to in British broadcasting since the 1960s (Tracey and Morrison, 1979: 112). Yet interestingly a recent web-based poll of people aged 15 and younger

showed that nearly half of them (46.7%) thought that there should be swear-words in young adult novels because normal teenagers swear and to represent them as not swearing was wrong. (URL: http://www.dream-tools.com/tools/polls.mv?view+youngadultspeech, accessed on 9 December 2002). See McEnery (2004) for an historical account of attempts to teach children not to swear.

15 It may be possible that this increase is not distributed smoothly across this period. This could lead to significant difference if we were able to subdivide this period further. However, on the basis of the available metadata we were not able to pursue this possibility further.

16 Category A is predicative negative adjective (e.g. *the film is shit*), M is imagery based on literal meaning (e.g. *kick shit out of*), R is 'reclaimed' usage with no negative intent, and T is religious oath used for emphasis (e.g. *by God*).

17 The annotation process involved one analyst annotating the corpus and a second analyst checking and verifying each annotation. Any cases of differences between the first and second analysis were resolved by both analysts discussing either similar past examples or, where the example was novel in some sense, deciding upon how such examples should be annotated in this case and in future.

18 There are 10 unclassifiable cases in the written section of the BNC, all of which are metalinguistic uses. In the spoken section, most instances of the X category are found in transcriptions where utterances are broken, i.e., the word following *fuck* is labelled as unclear.

19 It is our view that these examples represent either transcription errors in the corpus, or are faithfully transcribed disfluencies.

20 Other uses of *fuckers* are possible, e.g. *fucker and fuckee* (L) and *I haven't even sent off my fucker form yet* (E).

References

Andersson, L. and Trudgill, P. (1992) *Bad Language*. London: Penguin.

Aston, G. and Burnard, L. (1998) *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Collins, P. and Hollo, C. (2000) *English Grammar: An Introduction*. London: Macmillan.

Holmes, J. (1992) *An Introduction to Sociolinguistics*. London: Longman.

Hughes, G. (1991) *Swearing: A Social History of Foul Language, Oaths and Profanity in English*. London: Blackwell.

Jay, T. (1992) *Cursing in America: A Psycholinguistic Study of Dirty Language in the Courts, in the Movies, in the Schoolyards and on the Streets*. Philadelphia, PA: John Benjamins.

Kidman, A. (1993) 'How To Do Things with Four-letter Words: A Study of the Semantics of Swearing in Australia', unpublished BA thesis, University of New England.

Lakoff, R. (1975) *Language and Woman's Place*. New York: Harper & Row.

McEnery, A. (2004) *Swearing in English*. London: Routledge.

McEnery, A., Baker, P. and Hardie, A. (2000) 'Swearing and Abuse in Modern British English', in B. Lewandowska-Tomaszczyk and P. Melia (eds) *PALC'99: Practical Applications in Language Corpora*, pp. 37–48. Berlin: Peter Lang.

Montagu, A. (1973 [1967]) *The Anatomy of Swearing*, 2nd edn. London: MacMillan.

Nurmi, M. (1997) 'Swearing in America', available online at URL (accessed on 10 December 2002) http://www.uta.edu.fi/FAST/US1/P1/swear.html

Sheidlower, J. (1995) *The F Word*. New York: Random House.

Stenström, A. (1991) 'Expletives in the London–Lund corpus', in K. Aijmer and B. Alternberg (eds) *English Corpus Linguistics*, pp. 230–53. Harlow: Longman.

Tracey, M. and Morrison, D. (1979) *Whitehouse*. London: Macmillan.

Address

Anthony McEnery and Zhonghua Xiao, Department of Linguistics, Lancaster University, Lancaster LA1 4YT, UK. [emails: a.mcenery@lancaster.ac.uk; z.xiao@lancaster.ac.uk]