

## Unit 20 Swearing in modern British English (Case study 4)

### 20.1 Introduction

We discussed language variation in units 10.4 and 14 and sociolinguistics in unit 10.11. This case study explores variations in spoken and written registers in modern British English and demonstrates how corpora can be used in sociolinguistic studies. 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, Baker and Hardie 2000) has addressed the topic. Nonetheless, there is still no systematic account of swear words in English (though McEnery 2005 seeks to provide a better historical explanation of attitudes to bad language in English). In terms of methodology, swearing has been approached from the points of view of history (e.g. Hughes 1991), psycholinguistics (e.g. Jay 1992) and semantics (Kidman 1993). There have been, to date, few studies of swearing based on sociolinguistic variables such as gender, age and social class (see McEnery, Baker and Hardie 2000 for an exception). Such a study has been difficult in the absence of appropriate corpus resources. With the production of the British National Corpus (see unit 7.2), such a study became possible. In addition to parts-of-speech, the corpus is richly encoded with metadata pertaining to demographic features such as age, gender and social class, and textual features such as register, publication medium and domain. In this case study, we will explore such dimensions of variation to discover a general pattern of usage for one word, *FUCK*, in modern British English. 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*) or other matters, we decided to examine only the distribution pattern of *FUCK* (including 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 became more highly charged semantically, it has also acquired more grammatical flexibility so that *FUCK* ‘has altered from being exclusively a verb to every part of speech’ (Nurmi 1997).

For this study we will use BNCWeb. BNCWeb is a user-friendly interface to the BNC corpus. Note that the old BNCWeb query system is used in this study. Users of the BNCWeb World Edition query system (as used in case study 1) may obtain frequencies which are slightly different from those shown in the screenshots. This case study will introduce two other important features of BNCweb, namely distribution and cross-tabulation, and show you how to explore language variation using the metadata encoded in the corpus. Readers interested in a more comprehensive account of the use of *FUCK* in the BNC can refer to McEnery and Xiao (2004), on which this case study is based.

This unit consists of 5 sections. Unit 20.2 compares spoken and written registers. Unit 20.3 explores the pattern of *FUCK* usage in the spoken register while unit 20.4 explores the pattern of *FUCK* usage in the written register. Finally, unit 20.5 concludes.

### 20.2 Spoken vs. written register

The spoken register is generally more informal than the written register, and one of the linguistic indicators of informality is swearing (cf. Collins and Hollo 2000). In the BNC corpus, the spoken section consists of around 10% of the data while the other 90% are written texts. This section compares the distribution patterns of *FUCK* in spoken and written registers. To get the frequencies needed in this investigation, do the following:

1. Start *Internet Explorer*, type in the URL of BNCWeb and press the *Enter* key. You will be led to the website of BNCWeb, as shown in Fig. 20.1.
2. Click on the link *Log on the BNCWeb query system*, and you will be prompted to type in your user name and password (Fig. 20.2).
3. Enter your user name and password as required and confirm by pressing the *OK* button. Now the BNCWeb query system is ready for use, as shown in Fig. 20.3. You can explore the whole BNC corpus or select spoken or written texts alone.
4. As we are interested in comparing spoken and written registers, we will use the whole BNC corpus. But we will search for *FUCK* (including its morphological variants *fuck*, *fucked*, *fucks*, *fuckin(g)* and *fucker(s)*) separately so that we can have a clearer view of their distribution patterns across register. First type in *fuck* in the text box, select 1000 for the *Number of hits per page* and press the *Start query* button, as shown in Fig. 20.4.
5. Now you can see the concordance window for *fuck*. Click on the down arrow near *Thin* and select *Distribution* from the pull-down menu. Press the *Go* button (Fig. 20.5).
6. You will be taken to the *Distribution window* of *fuck* (Fig. 20.6). Record the *Number of words*, *Number of hits* and *Frequency per million words* for the spoken and written registers.
7. Now press the *Back* button on *Internet Explorer* a couple of times until you return to the interface of the BNCWeb query system (Fig. 20.3), and repeat steps 1-6 for the search strings *fucked*, *fucks*, *fucking|fuckin* and *fucker|fuckers* separately (the character | means *or*). Do the same for the search string *fuck|fucked|fucks|fucking|fuckin|fucker|fuckers* to find all of the instances of *FUCK*.

Your results should match those in Table 20.1. The normalized frequencies (NF) allow us to compare the distributions of individual word forms while word numbers and raw frequencies (RF) make it possible for us to calculate the log-likelihood score and significance level for the difference in frequencies by using SPSS statistics package (see case study 2).

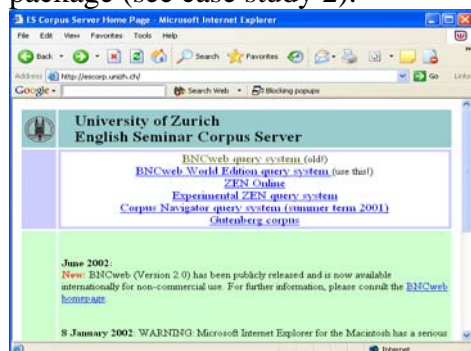


Fig. 20.1 The BNCWeb interface

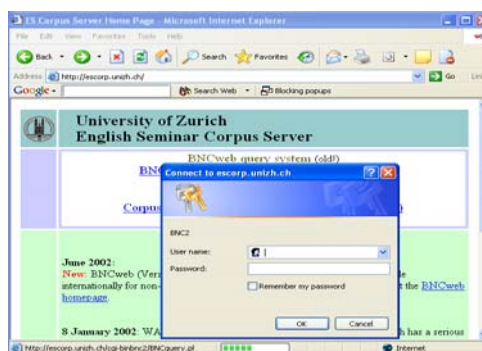


Fig. 20.2 Log on to BNCWeb

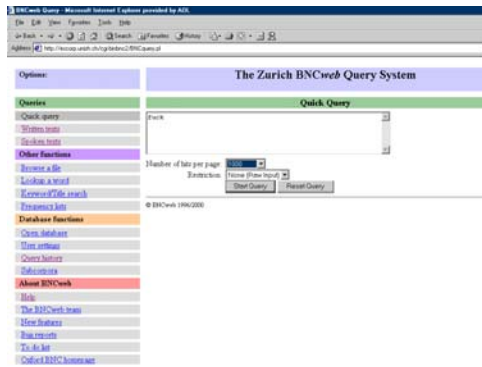
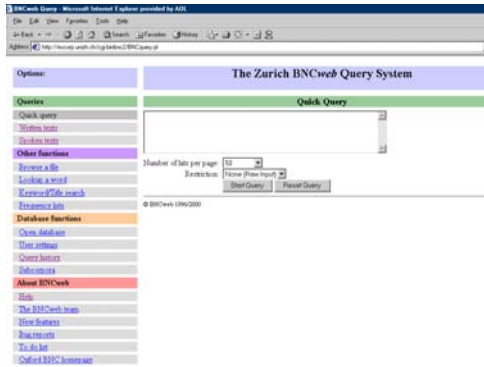


Fig. 20.3 The BNCWeb query system Fig. 20.4 Enter the search string

Your query "fuck" returned 1378 matches in 255 different texts. (freq: 13.77 instances per million words)

No.	Reference	Text	Context
1	AUS_15127	parents Arthur told them to go and	fuck
2	AUS_15124	for it, and whether parents should go and	fuck
3	AUS_651	I got as far as Chatterbox and I thought,	fuck
4	AUS_936	One day it's a revelation to say	fuck
5	AUS_1266	or is it her English from complexions, hand,	fuck
6	AUS_1293	skin, but that time I'm not going to	fuck
7	AUS_1464	Jay, that she says don't	fuck
8	AUS_15121	away," she squeals "Don't	fuck
9	AUS_1610	Jay had done your, and decide,	fuck
10	AUS_1730	your best? I'd happily proceed regularly	fuck
11	AUS_2278	I know you're scared of it,	fuck
12	AUS_2366	half drunk half my life, who goes in	fuck
13	AUS_2522	"OK, fuck your stupid Maths, and stepped off to the	fuck
14	AUS_2727	to maintain as usual, being a parent does you	fuck
15	AUS_2730	but totally remarkable to other people.	fuck
16	AUS_2878	Jay did (sneal, delighted)	fuck
17	AUS_3011	"no," said Jay "Why the	fuck
18	AUS_2551	Wow, I don't feel better, Lucy,	fuck
19	AUS_2986	pardon me for bringing you. It's just	fuck
20	AUS_7347	work not including. But I thought,	fuck

Your query "fuck" returned 1378 matches in 255 different texts. It was found in 255 texts.

Category	No. of words	No. of hits	Frequency per million words
Spoken	10,365,464	583	56.24
Written	8,974,0543	795	8.86
<b>total</b>	<b>100,106,007</b>	<b>1,378</b>	<b>13.77</b>

Fig. 20.5 The concordances for fuck Fig. 20.6 The distribution of fuck

Table 20.1 Spoken vs. written register

Form	Register	Words	RF	NF	LL ratio	Sig. level
fuck	Spoken	10365464	583	56.24	940.406	<0.001
	Written	89740543	795	8.86		
fucked	Spoken	10365464	62	5.98	68.066	<0.001
	Written	89740543	130	1.45		
fucks	Spoken	10365464	10	0.96	12.792	<0.001
	Written	89740543	18	0.2		
fucking	Spoken	10365464	2164	208.77	6150.587	<0.001
	Written	89740543	969	10.8		
fucker(s)	Spoken	10365464	25	2.41	28.841	<0.001
	Written	89740543	50	0.56		
All forms	Spoken	10365464	2844	274.37	6827.547	<0.001
	Written	89740543	1962	21.86		

As can be seen from the table, for all of the word forms under examination, the difference between spoken and written language is statistically significant at the level  $p < 0.001$ . *FUCK* occurs 12 times more frequently in spoken language than in written language. The greatest contrast is found for *fuckin(g)*, which was used nearly 20 times as frequently in the spoken section as in the written section of the corpus. While it is not clear why people use *FUCK* considerably more frequently in spoken language than in written language, our speculation is that *FUCK* occurs more frequently in informal than formal contexts. The censorship of published written texts is another possible explanation for the relatively low frequency of *FUCK* in written language. In spite of this quantitative difference, different word forms distribute across the registers in the same descending order: *fuckin(g)*, *fuck*, *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 unit is devoted to identifying these finer distinctions.

## 20.3 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 spoken data. We will also examine the possible influence of speaker gender, age, social class and education level on the pattern of uses of *FUCK*.

### 20.3.1 Demographically sampled vs. context-governed spoken language

As noted in unit 2.4, the BNC contains orthographically transcribed spoken language using two different sampling regimes: demographically determined and context-governed. To get the word numbers and frequencies of *FUCK* for the two types of spoken language, click on the down arrow on the right side of *General information* in step 6, select *Overall: Type of text* and press the *Show distribution* button, as shown in Fig. 20.7. You will be taken to the distribution window giving the word number, number of hits (RF) and frequency per million words (NF) for three types of text: spoken demographic, written and spoken context-governed (Fig. 20.8). Record the word numbers and frequencies for the two types of spoken text. Do the same for search strings *fucked*, *fucks*, *fucking*|*fuckin*, *fucker*|*fuckers*, and *fuck*|*fucked*|*fucks*|*fucking* |*fuckin*|*fucker*|*fuckers* as in step 7. Your results should match those in Table 20.2.

Spoken or Written	Category	No. of words	No. of hits	Frequency per million words
Spoken or Written:	Spoken	10,365,464	795	8.86
	Written	89,740,543	23	13.77
<b>Text Type (written):</b>				
Category	No. of words	No. of hits	Frequency per million words	
Magazines	19,664,309	277	27.34	
Arts	7,914,792	82	12.69	
Letters	8,991,740	32	5.89	
World affairs	15,243,340	24	2.23	
Applied science	7,341,375	24	1.91	
Social science	12,186,378	18	1.48	
Consumer and leisure	6,668,357	2	1.95	
Total and single	3,935,896	1	0.33	

Type of text:	Category	No. of words	No. of hits	Frequency per million words
Type of text:	spoken demographic	4,211,216	576	136.78
	written	89,740,543	23	8.86
	spoken context-governed	5,034,707	7	1.39
<b>total</b>		<b>98,986,466</b>	<b>1378</b>	<b>13.92</b>

Fig. 20.7 Show distribution

Fig. 20.8 Distribution across text type

Table 20.2 Spoken demographically sampled vs. spoken context-governed

Form	Type	Words	RF	NF	LL ratio	Sig. level
fuck	Demographic	4211216	576	136.78	838.609	<0.001
	Context-governed	5034707	7	1.39		
fucked	Demographic	4211216	61	14.49	86.922	<0.001
	Context-governed	5034707	1	0.2		
fucks	Demographic	4211216	10	2.37	15.729	<0.001
	Context-governed	5034707	0	0		
fucking	Demographic	4211216	2149	510.3	3218.681	<0.001
	Context-governed	5034707	15	2.98		
fucker(s)	Demographic	4211216	25	5.94	39.321	<0.001
	Context-governed	5034707	0	0		
All forms	Demographic	4211216	2821	669.88	4196.573	<0.001
	Context-governed	5034707	23	4.57		

With regard to the frequency of *FUCK*, the two types of spoken language differ significantly at the level  $p < 0.001$ . As can be seen in Table 20.2, demographically sampled spoken data contains 146 times as many instances of *FUCK* as context-governed spoken data. Some word forms, e.g. *fucks* and *fucker(s)*, are simply non-existent in context-governed spoken data, even though this part contains nearly one million more tokens than the first type of data.

Surprisingly, the contrast between the two types of spoken language is even more marked than the distinction between spoken and written registers. While context-governed spoken language is indeed more formal than demographically sampled spoken language (cf. Aston and Burnard 1998: 31), the difference between the two types of spoken language cannot be explained by the formal/informal distinction alone: written language is basically more formal than spoken language, yet the contrast between them is not as marked as that between two types of spoken language. A reasonable explanation is that the social contexts from which the second type of spoken data was sampled militated in favour of considerably fewer forms of *FUCK* than in the demographically sampled spoken data.

### 20.3.2 Gender of speaker

Men and women differ in their use of strong language (cf. Lakoff 1975: 5; Hughes 1991: 211; Holmes 1992: 171-176). For example, Stenström (1991) found from 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*. Consequently we decided to explore the hypothesis that the gender of speakers also influences the frequency of their use of *FUCK*.

Table 20.3 Gender of speaker

Form	Gender	Words	RF	NF	LL ratio	Sig. level
fuck	Male	4918075	337	68.52	50.025	<0.001
	Female	3255533	106	32.56		
fucked	Male	4918075	25	5.08	0.510	0.475
	Female	3255533	13	3.99		
fucks	Male	4918075	5	1.02	0.386	0.534
	Female	3255533	2	0.61		
fucking	Male	4918075	1394	283.44	353.624	<0.001
	Female	3255533	321	98.6		
fucker(s)	Male	4918075	18	3.66	8.967	0.003
	Female	3255533	2	0.61		
All forms	Male	4918075	1779	361.73	401.668	<0.001
	Female	3255533	444	136.38		

Table 20.4 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

Table 20.5 Comparison of the normalized frequencies across gender

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

To retrieve a range of word frequency data from the corpus related to male and female uses of *FUCK*, select *Speaker Gender* in the pull-down menu in the *Distribution* window in step 6.

Table 20.3 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 frequently 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 *fuckin(g)*, *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 (see McEnery and Xiao 2004 for a discussion of the semantic categories of *FUCK*).

On the other hand, while the use of *FUCK* differs quantitatively by speaker gender, it does not differ qualitatively. For both males and females, the rank and proportion of different word forms show a very similar distribution pattern (Table 20.4). Both genders use *fuckin(g)* most frequently, followed by *fuck*. While proportions of different word forms may vary slightly across gender, this variation is not statistically significant, as shown in Table 20.5.

### 20.3.3 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. (Holmes 1992: 183)

To test this hypothesis, we will first get the frequencies of *FUCK* used by speakers of different age groups by selecting *Speaker Age* in the *Distribution* window in step 6. Table 20.6 gives the frequencies of *FUCK* for different age groups. The table shows that for each of the forms of *FUCK*, and for all of word forms taken together, the difference in the distribution of *FUCK* across different age groups is statistically significant, though the significance level varies by word form, with the most marked contrast for *fuckin(g)*, followed by *fuck*. For all age groups, the most frequently used word form is *fuckin(g)*, 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, 25-34) appear to use *FUCK* more frequently than people of other age groups (Table 20.7). 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 aversion for the word. One plausible reason, in line with Holmes' hypothesis, for the relatively low frequency for age group 35-44 is that

parents with children and teenagers around them say *FUCK* less frequently than those who are yet to have children and those whose children have grown up and do not live with them. Children under the age of 15 use *FUCK* more frequently because they consciously want to shock adults and to behave in what they perceive to be an adult fashion. However, on the basis of corpus data alone, we cannot evaluate these explanations.

Table 20.6 Age of speaker

Form	Age	Words	RF	NF	LL ratio	Sig. level
fuck	0-14	460627	158	343.01	622.580	<0.001
	15-24	511858	126	246.16		
	25-34	1113709	93	83.50		
	35-44	1066857	8	7.50		
	45-59	1605978	46	28.64		
	60+	1122133	3	2.67		
fucked	0-14	460627	2	4.34	29.912	<0.001
	15-24	511858	10	19.54		
	25-34	1113709	5	4.49		
	35-44	1066857	1	0.94		
	45-59	1605978	2	1.25		
	60+	1122133	0	0		
fucks	0-14	460627	3	6.51	11.097	0.015
	15-24	511858	1	1.95		
	25-34	1113709	1	0.90		
	35-44	1066857	0	0		
	45-59	1605978	2	1.25		
	60+	1122133	0	0		
fucking	0-14	460627	217	471.10	1967.681	<0.001
	15-24	511858	638	1246.44		
	25-34	1113709	582	522.58		
	35-44	1066857	71	66.55		
	45-59	1605978	173	107.72		
	60+	1122133	18	16.04		
fucker(s)	0-14	460627	12	26.05	88.829	<0.001
	15-24	511858	18	35.17		
	25-34	1113709	8	7.18		
	35-44	1066857	0	0		
	45-59	1605978	0	0		
	60+	1122133	0	0		
All forms	0-14	460627	392	851.01	2613.071	<0.001
	15-24	511858	793	1549.26		
	25-34	1113709	689	618.65		
	35-44	1066857	80	74.99		
	45-59	1605978	223	138.86		
	60+	1122133	21	18.71		

Table 20.7 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 20.8 Cross-tabulation of speaker age and gender

Age	Gender	Words	RF	NF	LL ratio	Sig. level
0-14	Male	237530	248	1044.08	21.77	<0.001
	Female	223092	144	645.47		
15-24	Male	215310	657	3051.41	558.717	<0.001
	Female	296548	136	458.61		
25-24	Male	543791	643	1182.44	645.124	<0.001
	Female	569709	46	80.74		
35-44	Male	557551	64	114.79	26.657	<0.001
	Female	509306	16	31.42		
45-59	Male	531429	88	165.59	3.93	0.047
	Female	1072944	135	125.82		
60+	Male	531692	14	26.33	3.17	0.84
	Female	590441	7	11.86		

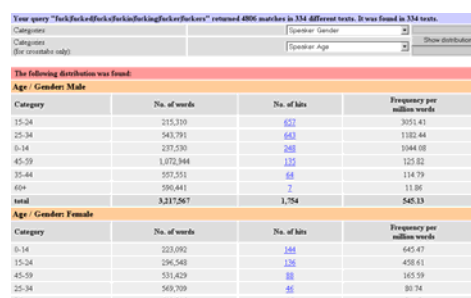


Fig. 20.9 The cross-tabulation of speaker gender and age

Yet if we cross-tabulate the variables speaker age and gender, a more distinct pattern can be observed. To do this, we need to select two variables in the *Distribution* window. Click on the first down arrow and select *Speaker Gender*. Click on the down arrow next to *no crosstabs* and select *Speaker Age*. Then press the *Show distribution* button, as shown in Fig. 20.9. Table 20.8 shows the result of the cross-tabulation. As can be seen from the table, 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 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.

### 20.3.4 Social class of speaker

The BNC classifies speakers into four social classes, namely AB, C1, C2 and DE. In this section, we will examine the possible influence of social class on the distribution pattern of *FUCK*. To get the frequencies of *FUCK*, select *Speaker: Social Class* from the pull-down menu in the *Distribution window* for each search string in step 6. Table 20.9 gives the frequencies of *FUCK* used by different social classes. As can be seen from the table, except for the word form *fucks* (there are only five instances of *fucks*, we doubt a statistical test based on such limited data can yield a reliable result), the difference in the distribution of all other word forms across social class is statistically significant. As with speaker gender and age, the greatest contrast is for *fuckin(g)*, 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.

Table 20.9 Speaker social class

Form	Class	Words	RF	NF	LL ratio	Sig. level
------	-------	-------	----	----	----------	------------



fuck	AB	696819	93	133.46	75.494	<0.001
	C1	427872	7	16.36		
	C2	485682	45	92.65		
	DE	267818	55	205.36		
fucked	AB	696819	18	25.83	15.993	0.001
	C1	427872	0	0		
	C2	485682	4	8.24		
	DE	267818	2	7.47		
fucks	AB	696819	3	4.31	1.987	0.583
	C1	427872	0	0		
	C2	485682	1	2.06		
	DE	267818	1	3.73		
fucking	AB	696819	187	268.36	297.527	<0.001
	C1	427872	39	91.15		
	C2	485682	305	627.98		
	DE	267818	198	739.31		
fucker(s)	AB	696819	1	1.44	8.087	0.012
	C1	427872	0	0		
	C2	485682	2	4.12		
	DE	267818	4	14.94		
All forms	AB	696819	302	433.4	339.734	<0.001
	C1	427872	46	107.51		
	C2	485682	357	735.05		
	DE	267818	260	970.81		

Table 20.10 Cross-tabulation of speaker gender and social class

Class	Gender	Words	RF	NF	LL ratio	Sig. level
AB	Male	266857	175	655.78	42.934	<0.001
	Female	413150	127	307.39		
C1	Male	187946	43	228.79	52.035	<0.001
	Female	239926	3	12.5		
C2	Male	169737	348	2050.23	654.976	<0.001
	Female	315945	9	28.49		
DE	Male	126512	176	1391.17	64.701	<0.001
	Female	138247	84	607.61		

The normalized frequencies for all forms show such a distinction. People from classes DE and C2 are most frequent users of *FUCK*, followed by AB. Interestingly, those from the class AB do not say *FUCK* less frequently than C1, especially people from age group 60+ (see Table 20.11). One might speculate that the older people from AB use *FUCK* frequently because they want to flaunt their seniority, while those from C1 show a considerably lower rate of *FUCK* usage because they consciously or unconsciously pay special attention to their linguistic behaviour so as to appear closer to how they perceive the AB speech to be. 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 hand (see Fig. 20.9 for cross-tabulation), as shown in Tables 20.10 and 20.11. Table 20.10 shows the result of cross-tabulation of gender and social class. As can 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 because both sexes from this class use *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.

Table 20.11 Cross-tabulation of speaker age and social class

Age	Class	Words	RF	NF	LL ratio	Sig. level
0-14	AB	127228	209	1642.72	24.550	<0.001
	C1	5722	0	0		
	C2	4439	1	225.28		
	DE	2	0	0		
15-24	AB	78210	80	1022.89	99.486	<0.001
	C1	40544	1	24.66		
	C2	29072	29	977.52		
	DE	42303	81	1914.76		
25-34	AB	101503	0	0	312.701	<0.001
	C1	55654	26	467.17		
	C2	192484	317	1646.89		
	DE	23468	4	170.44		
35-44	AB	81002	2	24.69	4.813	0.090
	C1	201306	17	84.45		
	C2	97480	10	102.59		
	DE	0	0	0		
45-59	AB	132275	0	0	431.876	<0.001
	C1	106972	2	18.7		
	C2	84611	0	0		
	DE	115857	168	1450.06		
60+	AB	94332	7	74.21	7.835	0.023
	C1	17674	0	0		
	C2	77596	0	0		
	DE	48244	0	0		

### 20.3.5 Education level of speaker

A common belief is that the better educated one is, the less likely one is to use bad language. A popular explanation for swearing is that people use swear words when they have few words at their disposal, i.e. their vocabulary is so impoverished that they have to use ‘easy’ and ‘lazy’ words in certain situations (cf. Andersson and Trudgill 1992: 65). This explanation is, in our view, unlikely to be true. The BNC encodes information pertaining to speaker’s education level, thus enabling us to test the influence of education on the use of *FUCK*.

Select *Speak: Education* in step 6 for the frequencies of *FUCK* used by speakers of different education levels, which are given in Table 20.12. Note that the table does not include the group *Still in education*. We decided to leave this group out of our discussion because this group may overlap with others. 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 do not use *FUCK* less frequently because they are mostly of the age group 15-24. As can be seen, people who left school at 15/16 are 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 – Young people are unlikely to leave school so early. 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, *fuckin(g)* is the most frequent word form, followed by *fuck* (see Table 20.13).

Table 20.12 Speaker education level

Education	Words	RF	NF	LL ratio	Sig. level
Left school 15/16	639039	596	932.57	762.703	<0.001
Left school 17/18	217282	32	147.27		
Educ. until 19/over	318267	16	50.27		
Left school 14/under	378669	9	23.77		

Table 20.13 Comparison of normalized frequencies across education level

Education level	Word form	NF	Rank
Left school 14/under	fucking	21.13	1
	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	$\frac{3}{4}$
	fucker(s)	6.26	$\frac{3}{4}$
	fucks	3.13	5
Left school 17/18	fucking	110.46	1
	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	---

## 20.4 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 will examine the possible influence of gender and age of author and audience, as well as the reception status of writing on the distribution pattern of *FUCK*.

### 20.4.1 Gender of author

We assume that author gender has a similar effect on the pattern of uses of *FUCK* to that of speaker gender. To test this assumption, we will first get the frequencies of *FUCK* used by male and female authors by selecting *Written: Gender of Author* in step 6. The results should match those given in Table 20.14. As can be seen from the table, male authors use *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 *fuckin(g)* demonstrating the greatest contrast. In terms of word forms, while female authors appear to prefer *fuck* to *fuckin(g)* more than male authors (see Table 20.16), 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 20.15). The fluctuation of the normalized frequencies can be discarded (LL=1.162, 3 d.f.).

Table 20.14 Gender of author

Form	Gender	Words	RF	NF	LL ratio	Sig. level
fuck	Male	31586324	486	15.39	28.625	<0.001
	Female	15497994	147	9.49		
fucked	Male	31586324	78	2.47	7.549	0.007
	Female	15497994	20	1.29		
fucks	Male	31586324	14	0.44	6.503	0.029
	Female	15497994	1	0.06		
fucking	Male	31586324	709	22.45	128.474	<0.001
	Female	15497994	132	8.52		
fucker(s)	Male	31586324	35	1.11	7.142	0.012
	Female	15497994	6	0.39		
All forms	Male	31586324	1322	41.85	162.124	<0.001
	Female	15497994	306	19.74		

Table 20.15 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 20.16 Comparison the normalized frequencies across gender

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

### 20.4.2 Age of author

Author age in written language is a sociolinguistic variable comparable to speaker age in spoken language and may, therefore, influence the distribution of *FUCK*. By selecting *Written: Age of Author* in the distribution window, you will get the word numbers and frequencies given in Table 20.17. 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 together. A comparison by word form shows that except for the two very infrequent words *fucks* (three instances) and *fucker(s)* (nine instances), all of the other word forms demonstrate a significant variation between age groups.

While young people also use *FUCK* a lot in written language as they do in spoken language, the pattern of using *FUCK* in written language appears to be different from that in spoken language in spite of some similarities, as shown in Table 20.18. 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 this

age group use *FUCK* more frequently than expected, though not as obtrusively as in spoken language. 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.

Table 20.17 Age of author

Form	Age	Words	RF	NF	LL ratio	Sig. level
fuck	0-14	581962	3	5.15	178.234	<0.001
	15-24	437149	3	6.86		
	25-34	1325516	97	73.18		
	35-44	2813226	32	11.37		
	45-59	2847335	36	12.64		
	60+	2451519	14	5.71		
fucked	0-14	581962	0	0	46.263	<0.001
	15-24	437149	0	0		
	25-34	1325516	20	15.09		
	35-44	2813226	5	1.78		
	45-59	2847335	11	3.86		
	60+	2451519	0	0		
fucks	0-14	581962	0	0	3.286	0.778
	15-24	437149	0	0		
	25-34	1325516	1	0.75		
	35-44	2813226	1	0.36		
	45-59	2847335	1	0.35		
	60+	2451519	0	0		
fucking	0-14	581962	12	20.62	121.236	<0.001
	15-24	437149	5	11.44		
	25-34	1325516	87	65.63		
	35-44	2813226	36	12.8		
	45-59	2847335	41	14.4		
	60+	2451519	21	8.57		
fucker(s)	0-14	581962	2	3.44	7.216	0.129
	15-24	437149	0	0		
	25-34	1325516	3	2.66		
	35-44	2813226	1	0.36		
	45-59	2847335	4	1.4		
	60+	2451519	1	0.41		
All forms	0-14	581962	17	29.21	336.394	<0.001
	15-24	437149	8	18.3		
	25-34	1325516	208	156.92		
	35-44	2813226	75	26.66		
	45-59	2847335	93	32.66		
	60+	2451519	36	14.68		

Table 20.18 Comparison of spoken and written languages

Age group	Spoken		Written	
	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

### 20.4.3 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 section, we will only consider the first three categories. Select *Written: Gender of Audience* in the distribution window. You will get the frequencies as given in Table 20.19. The table shows that 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 (see McEnery and Xiao 2004). Writing with an intended female audience contains significantly fewer occurrences of *fucked* than writings for an intended male audience. Other word forms (especially *fuck* and *fuckin(g)*) used for emphasis do not show a significant contrast.

Table 20.19 Gender of audience

Form	Gender	Words	RF	NF	LL ratio	Sig. level
<i>fuck</i>	Male	2451934	21	8.56	0.521	0.471
	Female	6235502	44	7.06		
	Mixed	54289029	591	10.89	---	---
<i>fucked</i>	Male	2451934	17	6.93	28.091	<0.001
	Female	6235502	3	0.48		
	Mixed	54289029	90	1.66	---	---
<i>fucks</i>	Male	2451934	0	0	---	---
	Female	6235502	0	0		
	Mixed	54289029	14	0.26	--	--
<i>fuckin(g)</i>	Male	2451934	24	9.79	1.405	0.236
	Female	6235502	45	7.22		
	Mixed	54289029	701	12.91	---	---
<i>fucker(s)</i>	Male	2451934	0	0	---	---
	Female	6235502	0	0		
	Mixed	54289029	43	0.79	---	---
All forms	Male	2451934	62	25.29	10.270	0.001
	Female	6235502	92	14.75		
	Mixed	54289029	1439	26.51	---	---

Interestingly, writing intended for a mixed audience is quite similar to writing intended for a male audience in terms of distribution patterns of *FUCK* (the difference is not statistically significant; LL=0.134, d.f.=1,  $p=0.714$ ) when all word forms are taken together. The difference in 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 *fuckin(g)*. 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.

### 20.4.4 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 will consider the first three categories, the frequencies of which can be obtained by selecting *Written: Age of Audience* in the distribution window.

Table 20.20 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 over 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 in *fuckin(g)*, 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 (there are only 2.73, 0.22 and 1.76 instances of *fucked*, *fucks* and *fucker(s)* per million words). This finding is in line with the social convention that writing for children avoids bad language more than writing for adults.

Table 20.20 Age of audience

Form	Age	Words	RF	NF	LL ratio	Sig. level
fuck	Adult	82335639	784	9.52	14.482	0.001
	Teenager	1697721	10	5.89		
	Child	969382	1	1.03		
fucked	Adult	82335639	128	1.55	0.755	0.712
	Teenager	1697721	2	1.18		
	Child	969382	0	0		
fucks	Adult	82335639	18	0.22	0.110	1.000
	Teenager	1697721	0	0		
	Child	969382	0	0		
fucking	Adult	82335639	960	11.66	22.217	<0.001
	Teenager	1697721	7	4.12		
	Child	969382	2	2.06		
fucker(s)	Adult	82335639	48	0.58	1.412	0.347
	Teenager	1697721	2	1.18		
	Child	969382	0	0		
All forms	Adult	82335639	1938	23.54	37.603	<0.001
	Teenager	1697721	21	12.37		
	Child	969382	3	3.09		

### 20.4.5 Reception status

In this section, we will examine the potential relationship between reception status and the pattern of usage of *FUCK*. The BNC classifies the reception statuses of written texts into four types: high, medium, low and unknown. We will discard cases where reception status is unknown. First select *Written: Reception Status* in the distribution window for each search string and get their frequencies. Your results should match those given in Table 20.21. As can be seen, 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 only significant for *fucks* and *fuckin(g)*.

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 20.22. The table by itself does not show a pattern of *FUCK* usage. However, if we combine Tables 20.21 and 20.22 and take statistical significance into consideration, we are able to see clearly the pattern of usage for *FUCK* across reception status.

Table 20.21 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 *High* and *Medium* in rows 1, 2 and 5 in Table 20.22 can be swapped, i.e. *High* (1), *Medium* (2) and *Low* (3). Note, however, that the ranks of *High* and *Medium* cannot be inverted for *fucks* and *fuckin(g)*, because the inverted order cannot explain the statistical significance as shown by *fucks* ( $p=0.007$ ) and *fuckin(g)* ( $p=0.012$ ). As the difference between high and medium reception statuses is significant for *fucks* and *fuckin(g)*, *High* and *Medium* cannot be swapped in rows 3 and 4. However, in row 3, *Medium* and *Low* can be swapped (i.e. *High* (1), *Medium* (2) and *Low* (3)) because the difference between these two categories is not statistically significant (LL=1.551, 1 d.f.,  $p=0.213$ ). These rearrangements clearly present the pattern of usage of *FUCK* across reception status: *High*>*Medium*>*Low*. This format is in harmony with the pattern observed when all word forms are taken as a whole, as shown in row 6 in Table 20.22. This finding is unusual but true. As such, swear words are very common in popular books and movies. The explanation for this phenomenon, however, is beyond the corpus-based approach and would require, at the very least, substantial sociological study to explain.

Table 20.21 Reception status

Form	Level	Words	RF	NF	LL ratio	Sig. level	LL ratio	Sig. level
fuck	High	24138350	278	11.52	1.353	0.245	73.179	<0.001
	Medium	31885282	402	12.61				
	Low	16488041	83	5.03	---	---		
fucked	High	24138350	40	1.66	0.776	0.381	8.456	0.015
	Medium	31885282	63	1.98				
	Low	16488041	15	0.91	---	---		
fucks	High	24138350	11	0.46	7.357	0.007	7.077	0.025
	Medium	31885282	3	0.09				
	Low	16488041	4	0.24	---	---		
fucking	High	24138350	402	16.65	6.252	0.012	179.914	<0.001
	Medium	31885282	447	14.02				
	Low	16488041	60	3.64	---	---		
fucker(s)	High	24138350	13	0.54	3.006	0.083	9.681	0.008
	Medium	31885282	30	0.94				
	Low	16488041	4	0.24	---	---		
All forms	High	24138350	744	30.82	0.639	0.424	245.785	<0.001
	Medium	31885282	945	29.64				
	Low	16488041	166	10.07	---	---		

Table 20.22 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

## 20.5 Unit summary and suggestions for further study

In this unit, we used the metadata information encoded in the BNC to explore the distribution pattern of *FUCK* both within and across spoken and written registers. While the investigation presented in this unit is only possible with appropriate corpus resources, we feel that the corpus-based approach is not all-powerful (cf. unit 10.15).



Corpora are useful in formulating and testing linguistic hypotheses, but they cannot provide explanations to questions such as ‘why do people from higher social classes use *FUCK* frequently?’. Nevertheless, the corpus methodology, in combination with other methodologies, is undoubtedly of use in providing descriptions that any purported explanations must account for. This unit gave you a step-by-step demonstration of how to use BNCWeb to explore language variation in the BNC. The BNC is extensively encoded with metadata. In addition to those factors encoded in the metadata explored in this unit, the distribution of *FUCK* may be influenced by many other factors encoded in the BNC. Amongst those you might care to examine *domain* of context-governed speech in the spoken register, as well as *date of creation* and *level of audience* in the written register.