Fuck revisited

Tony McEnery & Zhonghua Xiao, Lancaster University.

1. Introduction

This paper is a follow up to the investigation of McEnery, Baker and Hardie (2000) into the use of the word *fuck* in spoken British English. Both that paper and this are based on the British National Corpus. However, at the time of writing in 2000, the analysis of *fuck* in the written BNC had not been completed, hence the 2000 paper focussed on spoken English alone. In doing so, it explored the way *fuck* varied with respect to a range of meta-data encoded in the spoken BNC, principally age, sex and social class. We have now explored the written section of the BNC, and have explored the distribution of *fuck* with respect to a subset of the metadata encoded in the written BNC, namely domain, author gender, author age, audience gender, audience age, audience level, reception status, medium of text and date of creation. As some of these features have clear analogues in the spoken BNC (most clearly age and sex) comparisons between the work presented here and the earlier work on spoken English will be presented wherever possible. Throughout, unless otherwise stated, references to the frequency of usage of features in spoken language are taken from McEnery, Baker and Hardie (*ibid*).

2. Domain

This section examines the distribution of *fuck* in written language. Table 1 compares 9 written domains encoded in the BNC.

Table 1: Domains of written section of the BNC

Domain	Words	RF[1]	NF[2]	LL ratio	Sig. level
Imaginative	19664309	1485	75.52	2827.945	<0.001
Arts	7014792	208	29.65		
Leisure	8991740	98	10.9		
World affairs	15243340	73	4.79		
Commerce/business	6668357	29	4.35		
Social science	12186378	45	3.69		
Applied science	7341375	21	2.86		
Belief/thought	3035896	3	0.99		
Nat./pure science	3746901	0	0		

Clearly, the distribution of *fuck* is statistically significant by domain in written English. Forms of *fuck* are used most frequently in *imaginative writing*, probably because texts of this category are primarily fiction which contain a lot of reports of speech and are more akin to spoken language. 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 of *fuck*.

3. Gender of author

One might imagine that author gender would have a similar effect on the pattern of uses of *fuck* to that of speaker gender. In Baker, McEnery and Hardie (2000), males were found to use the word fuck much more frequently than females. This tendency does indeed seem to translate into writing also. As can be seen from Table 2, male authors use *fuck* more than twice as frequently as female authors. This difference is significant at the level of $p<0.001$ (LL=162.124, 1 df). 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*

[1] RF denotes the observed frequency of the feature in the BNC.
[2] NF denotes the normalised frequency per million words of the feature in the BNC.

more than male authors (see Table 4), the difference is not statistically significant (LL=0.439, 1 df). The proportion and rank of word forms show a very similar distribution pattern across author gender (Table 3). The fluctuation in the normalized frequencies can be discarded (LL=1.162, 3 df).

Table 2: 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 3: 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 4: Comparison the normalized frequencies of word forms 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				

4. 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*. Table 5 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. A comparison by word form shows that except for the two very infrequent words *fucks* (3 instances) and *fucker(s)* (9 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 6. In written English, the age group 60+ uses *fuck* least frequently. However, authors aged 25-24 use *fuck* most frequently, followed by the age group 45-59. While authors aged 45-59 use *fuck* slightly more often than those aged 34-45, the difference is not statistically significant (LL=1.721, p=0.217). Like speakers under 15, authors of the same age group use *fuck* are amongst the most frequent users of the word *fuck*, though the frequency of usage is not as high as in the 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 5: Age of speaker

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 6: 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

5. 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. Table 7 compares the use of different word forms across gender.

Table 7: Gender of audience

Form	Gender	Words	RF	NF	LL ratio	Sig. level
fuck	Male	2451934	21	8.56	0.521	0.471
	Female	6235502	44	7.06		
	Mixed	54289029	591	10.89	---	---
fucked	Male	2451934	17	6.93	28.091	<0.001
	Female	6235502	3	0.48		
	Mixed	54289029	90	1.66	---	---
fucks	Male	2451934	0	0	---	---
	Female	6235502	0	0		
	Mixed	54289029	14	0.26	--	--
fucking	Male	2451934	24	9.79	1.405	0.236
	Female	6235502	45	7.22		
	Mixed	54289029	701	12.91	---	---
fucker(s)	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	---	---

As can be seen from the table, when 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 less occurrences of *fucked* than writings for an intended male audience. Other word forms (especially *fuck* and *fucking*) used for emphasis 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, df=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 df). 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.

6. 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. Table 8 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 7 times as frequently as in writing for children. This difference is significant at the level of $p<0.001$. In terms of word forms, the greatest contrast is in *fucking*, followed by *fuck* while *fucked*, *fucks* and *fuckers* do not show a significant contrast because of the low overall frequencies of these word forms.[3] This finding is in line with the social convention that writing for children should avoid swearwords in an attempt to influence their linguistic behaviour, i.e. to encourage them not to use the words.[4]

[3] There are only 2.73, 0.22 and 1.76 instances of *fucked*, *fucks* and *fucker(s)* per million words.

[4] The desire to shield children from swearwords is apparent in other aspects of British public life, such as the decision not to broadcast such words on the television before 9 p.m. Such a decision has clearly had little influence on the linguistic habits of children, as shown in table 6, and teenagers themselves, it would appear, are not in favour of being shielded from such language - a recent web-based poll amongst children below the age of 15 showed that nearly half of them (46.7%) believed that there should be

Table 8: 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		

7. Level of audience

The BNC annotation scheme includes information pertaining to the levels of intended readership for a document, thus enabling us to explore the pattern of uses of *fuck* in this dimension. Table 9 compares the distribution of *fuck* in writings for different levels of audience.

Table 9: Level of audience

Form	Level	Words	RF	NF	LL ratio	Sig. level	LL ratio	Sig. level
fuck	Low	17126603	229	13.37	7.998	0.005	118.407	<0.001
	Medium	43837214	465	10.61				
	High	23967568	101	4.21	---	---		
fucked	Low	17126603	32	1.87	0.086	0.660	10.527	0.005
	Medium	43837214	77	1.76				
	High	23967568	21	0.88	---	---		
fucks	Low	17126603	5	0.29	0.384	0.826	0.853	0.671
	Medium	43837214	9	0.21				
	High	23967568	4	0.17	---	---		
fucking	Low	17126603	243	14.19	2.73	0.098	52.212	<0.001
	Medium	43837214	547	12.48				
	High	23967568	179	7.47	---	---		
fucker(s)	Low	17126603	13	0.76	0.001	0.980	12.749	0.002
	Medium	43837214	33	0.75				
	High	23967568	4	0.17	---	---		
All forms	Low	17126603	522	30.48	9.711	0.002	178.857	<0.001
	Medium	43837214	1131	25.8				
	High	23967568	309	12.89	---	---		

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 is statistically significant for all words expect *fucks*, which occurs rarely. The greatest contrast is found for *fuck* (LL=118.407). It is also interesting to note that the medium level is closer to the low level than to the high level. Except for *fuck*, the difference between different audience levels is not

---

swearwords in young adult novels because normal teenagers swear and to avoid using swearwords would be unnatural. (http://www.dream-tools.com/tools/polls.mv?view+youngadultspeech, accessed on 9th December 2002).

statistically significant. While it is not clear why *fuck* shows a significant contrast, we speculate that the contrast 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 df). But this significance level is weighted by the marked contrast for *fuck*.

8. Reception status

In this section, we will 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 will discard cases where reception status is unknown. As can been seen from Table 10, 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 *fucking*.

Table 10: 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	---	---		

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 11. The table by itself does not show a pattern of *fuck* usage. However, if we combine Tables 10and 11 and take statistical significance into consideration, we are able to see clearly the pattern of usage for *fuck* across reception status.

Table 11: 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

Table 10 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 11 can be swapped, i.e. *High (1)*, *Medium (2)* and *Low (3)*. Note, however, that the ranks of *High* and *Medium* cannot be inverted, because the inverted order cannot explain the statistical significance as shown by *fucks* (p=0.007) and *fucking* (p=0.012). As the difference between high and medium reception statuses is significant for *fucks* and *fucking*, *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 df, 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 11. This finding is unusual but true. The explanation for this phenomenon, however, is beyond the corpus-based approach and would require, at the very least, substantial sociological study to explain.

9. Medium of text

Five basic types of medium of text are annotated in the BNC corpus. This section will use this information to examine the effect of publication medium on the distribution pattern of *fuck*. Table 12 compares the rate of usage of *fuck* across medium.

It is clear that for all of the word forms, the contrast between types of medium is statistically significant. 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* can be re-ordered. *Book* ranks before *miscellaneous* only for *fucking*. Yet the difference in its frequencies between the two types of medium is significant, therefore *book* and *miscellaneous unpublished* cannot be re-ordered. *Fuck* is most frequently used in *book*, followed by *miscellaneous unpublished*, *periodical*, *miscellaneous published* and *written-to-be-spoken*. As can be seen from the table, *fuck* occurs nearly 5 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.

Table 12: Medium of text

Form	Medium	Words	RF	NF	LL ratio	Sig. level	LL ratio	Sig. level
fuck	Book	52574506	667	12.69	0.198	0.657	265.830	<0.001
	Mis. unpub.	3461953	47	13.58				
	Periodical	23978695	80	3.34	---	---		
	Mis. pub.	3922977	1	0.25	---	---		
	To-be-spoken	861592	0	0	---	---		
fucked	Book	52574506	100	1.9	0.740	0.390	22.373	<0.001
	Mis. unpub.	3461953	9	2.6				
	Periodical	23978695	19	0.79	---	---		
	Mis. pub.	3922977	1	0.25	---	---		
	To-be-spoken	861592	0	0	---	---		
fucks	Book	52574506	16	0.3	0.619	1.000	11.720	0.014
	Mis. unpub.	3461953	2	0.58				
	Periodical	23978695	0	0	---	---		
	Mis. pub.	3922977	0	0	---	---		
	To-be-spoken	861592	0	0	---	---		
fucking	Book	52574506	875	16.64	22.333	<0.001	430/306	<0.001
	Mis. unpub.	3461953	25	7.22				
	Periodical	23978695	59	2.46	---	---		
	Mis. pub.	3922977	8	2.04	---	---		
	To-be-spoken	861592	0	0	---	---		
fucker(s)	Book	52574506	41	0.78	0.030	1.000	11.007	0.018
	Mis. unpub.	3461953	3	0.87				
	Periodical	23978695	6	0.25	---	---		
	Mis. pub.	3922977	0	0	---	---		
	To-be-spoken	861592	0	0	---	---		
All forms	Book	52574506	1699	32.32	6.137	0.013	709.749	<0.001
	Mis. unpub.	3461953	86	24.84				
	Periodical	23978695	164	6.84	---	---		
	Mis. pub.	3922977	10	2.55	---	---		
	To-be-spoken	861592	0	0	---	---		

10. Date of creation

In this section, we will compare written English in the periods 1960-1974 and 1975-1993 to see whether language change has influenced the pattern of uses of *fuck* in British English. As date of creation is encoded for the written section of the BNC alone, it was not possible to examine changes in the distribution pattern of *fuck* in spoken English in McEnery et al (2000). 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 13: Date of creation

Form	Date	Words	RF	NF	LL ratio	Sig. level
fuck	1975-1993	75501632	762	10.09	5.241	0.022
	1960-1974	2036939	11	5.4		
fucked	1975-1993	75501632	128	1.7	6.815	0.009
	1960-1974	2036939	0	0		
fucks	1975-1993	75501632	18	0.24	0.958	1.000
	1960-1974	2036939	0	0		
fucking	1975-1993	75501632	937	12.41	0.020	0.888
	1960-1974	2036939	26	12.76		
fucker(s)	1975-1993	75501632	47	0.62	1.642	0.200
	1960-1974	2036939	3	1.47		
All forms	1975-1993	75501632	1892	25.06	2.520	0.112
	1960-1974	2036939	40	19.64		

As can be seen in Table 13, when all word forms are taken together, there is no significant difference in the frequency of *fuck* between the two periods under consideration, in spite of a 5% increase in 1975-1993. 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 *fuck* doubled in the latter period. The difference in the frequencies of *fucker(s)* is not significant, but the use of the word was reduced by half in 1975-1993. It is also interesting to note that, *fucked* and *fucks* appear to be a new development in 1975-1993, because the two-million-word texts sampled for 1960-1975 does not contain a single instance of the two words

11. Conclusion

Fuck in written English generally acts as we would expect it to, given how it acts in spoken English. It is correlated with writing for a lower level of audience, as it is associated with speech from the lower classes. It is a marker of male readership/authorship as it is a marker of male speakers. Also, it is a word used more frequently by younger writers, just as it is a word more often spoken by younger speakers.

However, the written BNC also allows us to explore the influence of context of use on the word more clearly than we can using the spoken BNC. As one would expect, the word is associated with more informal sorts of writing, being totally absent from writing in the natural and pure sciences, but most frequent in imaginative writing. The issue of hearer, which is vexatious to explore with the spoken BNC (see McEnery, Baker, Hardie 2000) is translated into intended reader in the BNC. This research question is more tractable, and we can see that texts intended for females shun the word relative to texts intended for males. The effect of this avoidance of *fuck* is even noticeable in texts intended for mixed audiences, which are more akin to texts for female audiences than male with regard to the frequency of their usage of the word *fuck*. Yet running contrary to these findings is the finding related to reception status, where high reception status is linked to an elevated level of *fuck* usage.

While the investigation presented in this paper is only possible with appropriate corpus resources, we feel that the corpus-based approach is not all-powerful (cf. McEnery, Baker & Hardie 2000: 47) and often the process of explaining corpus findings leads one to methodologies other than the corpus method (McEnery, 2003). Corpora are useful in formulating and testing linguistic hypotheses, but they cannot readily provide explanations for questions such as why people from higher social class use *fuck* frequently and why writings of the highest reception status contain the most frequent use of *fuck*. Nevertheless, the corpus

methodology, in combination with other methodologies, is undoubtedly of use in providing descriptions that any purported explanations must account for.

References

McEnery, A. 2003. *Swearing in English*. London: Routledge.

McEnery, A, Baker, P & Hardie, A. 2000. Swearing and abuse in modern British English. In Lewandowska-Tomaszczyk, B & Melia, P (eds.) *PALC'99: Practical Applications in Language Corpora*. : Europäischer Verlag der Wissenschaften: Peter Lang.