

## AUTOMATIC RETRIEVAL AND THE FORMALIZATION OF MULTI WORDS EXPRESSIONS WITH F-WORDS IN THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH

*Prihantoro\**

### ABSTRACT

*Penelitian ini bertujuan untuk mengungkapkan 1) peran leksikogrammar dalam menentukan kutub polaritas Kata-F dan 2) bagaimana cara formalisasinya untuk menunjang pengolahan korpus. Data diperoleh dari Contemporary American English Corpus (COCA). Setelah dibandingkan dengan korpus bahasa Inggris lain korpus ini memiliki frekwensi kata-F tertinggi. Penulis mengaplikasikan metode analisis korpus dengan mengirimkan query untuk menelusuri data kata-F pada antarmuka COCA. Hasil penelusuran menunjukkan bentuk lingual frasa dan klausa yang mendampingi F-word dapat menunjukkan polaritas kata-F, yang sebetulnya tidak selalu negatif. Penulis juga mendesain computational resource yang bisa digunakan secara off-line untuk menelusuri teks digital yang dapat dikumpulkan secara mandiri, sehingga tidak harus menggunakan korpus yang hanya bisa diakses online.*

**Kata kunci:** kata-F, korpus, penelusuran informasi, Bahasa Inggris Amerika, polaritas

### ABSTRACT

The research problems in this research are 1) how lexicogrammar takes role in determining polarity of F-Word<sup>1</sup> and 2) how to formalize it for corpus processing. The data is obtained from the Contemporary American English Corpus (COCA). In this corpus, F-word is proven to be highest in frequency as compared to its distribution across corpora. Corpus methodology is applied by sending queries to retrieve F-Words to COCA interface. Tokens combination surrounding F-words resulted in the phrase and clause unit accompanying F-words, which are significant cues to determine F-word polarity. The polarity is later proven to be not necessarily negative. I also designed a computational resource to allow the retrieval of F-words offline so that users might apply it to any digital text collections.

**Keywords:** F-words, Corpus, Information Retrieval, American English, Polarity

---

\* Faculty of Humanities Diponegoro University Semarang, Indonesia

## INTRODUCTION

This section introduces the background of why F-word becomes the focus of the study. Following the background is previous studies, and the description of how this research differs from them. Data in this section is focused on F-words obtained by querying; however, phrases and clauses as immediate context, shown by concordances, are important cues so that they are included in the analysis and will be referred as multiword units (MWUs), terms commonly used in corpus and computational linguistic tradition. The final part of this section emphasizes in the procedure of data collection and analysis.

Deciding a corpus to research is an important preliminary step in a corpus study. A corpus can be built specifically for a particular study, such as the study by Nadar, et al (2005:1)<sup>2</sup>. The corpus used in the kind of Nadar's study is usually specialized in domain, small in size and purposive in content. Another type is general corpora that are larger in size and more various in the domains, where the content is representatively distributed over periods of time and varieties of genres. When already established, this kind of corpus is usually shared online, allowing users to research its content from various angles.

Besides the type, what matters in determining which corpus to use in a research is often the adequacy of data, which can be obtained by understanding the frequency distribution across corpora. In the findings and discussion section, I will show the distribution of F-Words across English varieties by the frequency identification of F-words over corpora of world English, which is obtained from [www.americancorpus.org](http://www.americancorpus.org)<sup>3</sup>. It is discovered, and later be discussed in more detail, that the highest frequency of F-words are in American English as represented by the Corpus of Contemporary American English. This is one reason for the author to use the corpus in this research. Another reason is the corpus standard. Davies (2008:2) believes that COCA is a large, representative, and well balanced corpus that is freely available online.

Subjects of this research are words that are often censored like 'fuck', 'fucker', 'fucking' which will be referred as <fuck> or F-words henceforth. In several standard dictionaries

of English <fuck> is a lemma with multiple categorizations of Part of Speech (POS) and many different senses. It can be a noun, a verb, adjective, adverb and interjection. It can be used to show anger (fuck you!), pain (fuck! that hurts!), apathy (I don't give a fuck to what he said!) and etc. One of the reasons for the multiple interpretation of <fuck> is the nature of its companion words (collocates). This follows the view of Firth's functional sentence perspective (1957:82), who believes that certain words are more likely to occur with particular words as compare to other words. For this reason, understanding the collocates of <fuck> is one crucial importance.

Studying syntax<sup>4</sup> query of F-word is another crucial importance, since it is used in the technology of information retrieval (internet, natural language processing, corpus linguistics etc). This word appears in mostly non-formal setting and its writing does not always seem to follow English writing convention. The variants might escape from dictionary record, but they are present in the corpus.

Literatures regarding F-words are often related to X-phemism (Orthopeism, Euphemism, and Dysphemism). Orthopeism refers to a word used to describe the reference literally. When the referent is a taboo subject, euphemism is used to 'soften' the language. Dysphemism, on the other hand, is the opposite of euphemism, where the word that is considered 'rude' or 'taboo' is used. See Allan & Burridge (2006:32)

The participants and setting of the talks might determine which X-phemism to use. One of the examples is, using the word 'faeces', instead of 'poo' to biologists or doctors. Biologists and doctors are most likely to understand this term. In a seminar or academic journal writing, this term can also be considered appropriate. When speaking to kids, using 'poo' might be considered proper. However, using 'poo' in a biology journal is inappropriate.

It is commonly believed that euphemism and Orthopeism are preferred over dysphemism. Orthopeism is formal and direct, while euphemism is more colloquial and figurative (Allan & Burridge, 2006: 44). The dispreferred one is dysphemism as it is considered harsh or offensive. This might be one of the reason why in the history of dictionary

ensorship, <fuck> has undergone censorship several times.

Lemma <fuck> has been censored few times to appear in dictionary entry. This makes tracing its etymology difficult. Its first appearance in dictionary was censored. This dates back to when Samuel Johnson's English Dictionary was published in 1755 (in Hughes, 2006 and Leigh & Lepine, 2005). Webster Dictionary also censored this entry in the three editions (editions published in 1806, 1828 and 1961). It appeared in Penguin English Dictionary published in 1965 edition and in the supplement of Oxford English Dictionary published in 1975. See Hughes (2006:29). These days, in contemporary English dictionaries, <fuck> is already one official entry, marked by uppercased caution RUDE or HARSH.

The works of Hobbs (2012), Norrick (2012) and Kravis (2013) study f-words from literary perspectives. Other literatures of F-words are related to feminism, such as the work of Fairman (2007), Olive & Thorpe (2011), Jaworska & Krishnamurthy (2012), and Harding et al (2013). Busthomi (2010) studies how F-word is used as an intimate marker. The work of Jaworska and Krishnamurthy (2012) focused on the corpus in Anglo American setting. In linguistics, F-words are often widely discussed in the literatures regarding to swearwords, such as the works of Brewster (2013), Leigh & Lepine (2005), McEnery, Baker, & Hardie (2000), McEnery & Xiao (2004), Howe (2012). The last three works focused on the use of f-words in British National Corpus (BNC). There are also some specific works about f-words. Sheidlehower (2009) documented massive entries and sub entries of F-words, particularly compounds. The retrieval of F-words in global web corpus (shown in the beginning of the finding section) indicates that f-words are frequently used in American English as compared to other English varieties. His work, however, is not corpus based. Therefore, a serious research on American English corpus is a crucial importance. One that is quite similar to this study is the work of Chiorean (2014). It studies the distribution of F-words across COCA domains. However, this study did not focus on what words that may accompany the F-words and its morpho-syntactic configuration in formulating Multi Word Units (MWUs). It does not also offer formalization

so that the resource can be used on-line.

My study here seeks to describe the collocates of F-words, and its lexicogrammatical configuration. It also seeks answers to question whether it is possible that F-words has positive value. Unlike the previous works, I formalized the research outcome. Therefore, users can also apply the resource to other corpora or a corpora of his/her own creation that is not published online.

Representativeness is one important issue in Corpus linguistics. Different variables are proposed, but several features are always shared such as the size and variety. See Mc Enery & Hardie (2012:21), O'Keefe & McCarthy (2010:55), and Ludeling & Kytö (2009:11). The data in this research is obtained from corpora collection maintained by Mark Davies, a professor in Brigham Young University. The corpora collection is mostly comprised of English corpora, although Spanish and Portuguese corpora are also available (<http://corpus.byu.edu/>).

Corpus of contemporary American English (COCA), one of the corpora, is comprised of different sections (TV-shows, magazines, newspapers, etc). This will make the research of F-Words be comprehensive and be proportionally stratified over the genres of where the F-words are used. However, before discussing the nature of F-words in COCA, it is also important to understand how this research is carried out and how the data is presented in this paper

Although data from other English corpora (historical, British and Canadian English corpora) are also consulted as secondary source, this research focuses on the use of <fuck> in COCA. The corpora are explored by using regular expression that follows the syntax of COCA. All queries are performed by regular expression-like method, but the presentation of the results may vary. The procedure of data collection and analysis are summarized as follow:

- 1) Sending F-word queries to diverse English corpora<sup>5</sup> to understand which corpus has the highest frequency of F-words
- 2) Comparing frequencies across corpora and selecting a corpus with the highest frequency (which is COCA).
- 3) Sending F-word queries to COCA<sup>6</sup>

- interface.
- 4) Obtaining supporting details such as frequency, domains, and historical chart.
  - 5) Displaying concordance to understand how the queries are used in left and right hand side contexts. The focus of this research is on word, but meaning is often contextualized by immediate tokens which structure Multi Word Units (MWUs) in terms of phrases or clauses.
  - 6) Investigating whether the lexicogrammar of MWUs (phrase and clause) is parallel to positive or negative polarity. The method is bottom-up, which is by retrospectively identifying the polarity and examining recurrent patterns. When necessary, supporting details and extended context (discourse unit) is also consulted.
  - 7) Using the lexicogrammar features as a guide to build a computational resource that can be used to explore off-line digital texts, so that researchers can explore his/her self customized/created corpus.

The initial result presentation is isolated query result, which shows the result without the presence of the context. For instance, the query that aims on retrieving 'fuck' plus any word will result in sequences like 'fuck you' 'fuck off', 'fuck up' and etc with frequency hits. Note that only these sequences are displayed without left or right context. When full concordance is required, one of the results can be selected. This will display concordance with left and right context, also known as Keyword(s) in Context (KWIC), with metadata information (source of publication). When more contextual cues are required, a look on the extended context (file view) is also possible.

Besides concordance, the result of the retrieval can also be displayed by using two charts: periodical and sections charts. Periodical chart shows the frequency of occurrence over years, while section charts shows the frequency distribution over different sections (spoken, academic, fiction etc).

The results of the retrieval in COCA are categorized into sequences of Multi Word Units

(henceforth, MWUs). These sequences are described by using Local Grammar Graph (LGG) formalism, a finite state automaton based graph that is exploitable in Unitex, a corpus processing software. Some linguistic and computational issues related to the retrieval of F-words are also discussed. Another part of the discussion is related to the polarity of F-words, which according to the dictionary meaning is inherently considered negative. I will show several lexico-grammatical configurations that can make the polarity positive.

## DISCUSSION

This findings section begins with the distribution of F-Words across English Varieties. Subsequently, the discussion is focused on COCA data. The procedures of data analysis is explored in detail here; which is how to obtain data, and used immediate contexts as cues<sup>7</sup> in determining positive or negative polarity. I will also show how extended contexts and extra linguistic factors take significant role in making decision about the polarity. The MWU's based computational resource design is also described afterward by considering orthographical challenges.

### F-Words Distribution across English Corpora

The distribution of so-called 'rude' words may differ even within a language. For instance, the word 'bugger' is an insult that is used in British English, but not quite frequent in American or Canadian English. Figure 1 presents us with the distribution of F-words in different variety of world Englishes (*see endnotes*).

In figure 1, the highest frequency of occurrence of F-words is observed in US. Although significant occurrences are also observed in GB (British English) and CA (Canadian English), it is lower than American English. As Global Web English Corpus looks at the distribution of English used in different parts of the world, it is also important to understand the distribution of F-words in specific English variety. Consider the retrieval of <fuck> in BNC, a corpus of British English, as shown by figure 2 (*see endnotes*).

Figure 2 shows the result of the retrieval for lemma <fuck> showing different inflectional and derivational forms. The highest frequency

of occurrence is observed for ‘fucking’ <3039>. The distribution, as English speakers are aware, is supposed to be in spoken section. However, let us validate this by looking at the distribution chart as shown by figure 3 (see endnotes).

Figure 3 validates the hypothesis that F-words are used in spoken section instead of the other sections. Fiction ranks the second. Now, let us consider the distribution of F-words in Strachy Corpus, a Canadian English Corpus. See figure 4 (see endnotes).

The frequency of occurrence of F-words in Canadian English is much lower <127>. However, unlike the BNC, the highest frequency of occurrence is on the fiction section. The frequency of occurrence in spoken section is only <14> time. See figure 5 (see endnotes).

The distribution of F-Words in BNC and Canadian English Corpus are different in terms of the section of occurrence. In BNC, the frequency is observed high in spoken section. The fiction section ranks the second. While in Canadian English Corpus, the first rank is fiction and the second one is magazine section. The frequency is also lower in Canadian English. However, in terms of the rank of word forms, the first rank of

occurrence is shared. The word form ‘fucking’ ranks first in both BNC and Canadian English Corpus. The second rank is also shared, which is ‘fuck’. Now, let us have a look at the occurrence and the distribution in COCA. See figure 6 (see endnotes).

Unlike BNC and Canadian corpus, ‘fuck’ ranks the first, and ‘fucking’ ranks the second in COCA. A significant gap with the previous two corpora is observed in terms of frequency of occurrence. The frequency of swearing is very high (7936 for ‘fuck’ and 7483 for ‘fucking’). That the frequency of occurrence in American English corpus is highest than other English varieties also positively correspond to the frequency of occurrence of lemma <fuck> in global web English corpus

### The Configuration of F-Words in COCA

In the website (<http://corpus.byu.edu/coca/>), the syntax query is fully documented. Several metasympol are used in the queries to optimize the outcome of the retrieval. Result in table 2 is the comparison of several different queries performed in COCA:

**Table 1.**  
**Retrieving Word Forms with Different Queries**

	[fuck]		*fuck		fuck*		*fuck*	
1	fuck	7909	fuck	7936	fuck	7936	fuck	7936
2	fucked	1247	motherfuck	15	fucking	7483	fucking	7483
3	fucks	234	dumbfuck	15	fuckin	2102	fuckin	2102
4	fucking	183	clusterfuck	12	fucked	1366	fucked	1366
5			bumfuck	8	fucker	376	motherfucker	602
6			mind-fuck	7	fucks	235	fucker	376
7			-fuck	6	fuckers	146	fucks	235
8			what-the-fuck	5	fucked-up	131	motherfuckers	151
9			rat-fuck	5	fuck-up	46	fuckers	146
10			ratfuck	4	fucken	41	fucked-up	131

There are four methods, which I used to retrieve word forms of lemma <fuck>. The first one is using brackets. The result, however, only retrieved the inflected forms. It missed identifying the derivation such as ‘fucker’ and ‘fuckers’. Using star (\*) means to retrieve any character(s) that are attached to ‘fuck’ that is not separated by space such as, ‘motherfucker’, ‘clusterfuck’ or ‘ratfuck’. There are also words that are unconventionally written such as ‘fucken’, ‘fuckin’ and ‘-fuck’. This orthographical issue might seem trivial, but very important in terms of information retrieval.

That F-word can take the form of single word or sequence of multiword units (MWUs) is observed from table 2. When taking the form as a single word, it might fall to one of classical POSs categorization such as noun, verb, adverb, interjection or adjective. Consider the concordance of ‘fuck’ in figure 7 (see endnotes).

Unlike previous outcome, which was context-less, the concordance may show how the words are used in context. In figure 8, the full concordance shows concordance lines where not all of them fall within one POS category. Notice in the second line, ‘fuck’ is an interjection (indicated by!), showing pain. In line three, ‘fuck’ is a verb (followed by an

object noun) showing that the speaker is annoyed, while in line 4 fuck is a noun, which seems to refer to god ( a subject before the verb knows). At this point, collocate (companion word) is very important in determining the sense.

### Collocation Sequences

I analyzed the collocation both right and left context. The top 100 collocates are chosen for each query result. The results that include punctuation are removed. When recurrent sequences are observed, they are documented and formalized. The formalization is performed by UNITEX graph editor to make it directly exploitable for computational purpose.

### Verb

The verb form of <fuck> ‘to copulate/to have sex’ can be both transitive and intransitive. Even though different type of nouns may take the position of object for transitive verb <fuck>, but the most frequent type is pronoun, specifically ‘you’. See the frequency as compared to other pronouns in table 2:

**Table 2.**  
**The Collocation of Pronouns with <fuck.v>**

FUCK YOU	1159	FUCKING ME	19	FUCKED THEM	8	FUCK ANYBODY	3	FUCKS EVERYBODY	1
FUCK IT	319	FUCK US	19	FUCK MYSELF	8	FUCK HERSELF	3	FUCKING YOURSELF	1
FUCK ME	208	FUCK THEY	17	FUCK EVERYONE	8	FUCK YOUS	3	FUCKING ONE	1
FUCK HER	115	FUCKED YOU	16	FUCK ANYTHING	7	FUCK YOURSELVES	3	FUCKING HIMSELF	1
FUCK HIM	106	FUCK EVERYTHING	13	FUCKING EACH	7	FUCK OURSELVES	3	FUCKED YOURS	1
FUCK YOURSELF	80	FUCKING HIM	12	FUCKED EACH	6	FUCK ONE	3	FUCKED SOMETHING	1
FUCKED HER	61	FUCK WHO	11	FUCK EVERYBODY	6	FUCK SOMETHING	2	FUCKED YA	1
FUCK THEM	57	FUCK WE	11	FUCK EACH	6	FUCKED HERSELF	2	FUCKED WHO	1
FUCKED ME	45	FUCK THEMSELVES	11	FUCK YA	5	FUCKED SOMEONE	2	FUCKED WE	1
FUCK I	38	FUCK SHE	10	FUCK SOMEBODY	5	FUCKS IT	2	FUCKED SOMEBODY	1
FUCKING HER	29	FUCKED US	10	FUCK SOMEONE	4	FUCKS THEM	2	FUCKED MYSELF	1
FUCK HIMSELF	28	FUCKS YOU	9	FUCK ITSELF	4	FUCKS WHO	2	FUCKED HE	1
FUCKED HIM	25	FUCKING THEM	9	FUCKING US	4	FUCKS THEY	1	FUCKED ANYBODY	1
FUCK HE	23	FUCKS ME	8	FUCKED HIMSELF	3	FUCKS SOMEONE	1	FUCK WHOEVER	1
FUCKED IT	20	FUCKS HER	8	FUCKED EVERYTHING	3	FUCKS HERSELF	1	FUCK PLENTY	1

In table 2, the frequency of ‘fuck you’ outranks the others <fuck><PRO> sequence. At this point, it is important to understand that COCA retrieves the same expressions as the query regardless of the meaning. The sense of ‘fuck you’ does not necessarily correspond to sex. See example 1 and 2:

- (1) I fuck you real good, baby “, etc.... ON THE CORNER: A WHORE pretends not
- (2) I’m gay, and if you don’t like it, fuck you.

In (1), <fuck.V> is used in the sense of having sex, but not in (2). In (2), it shows that the speaker is annoyed with the addressee who does not like gay. Now have a look at the concordance of other pronouns as presented by (3) to (6):

- (3) Then I asked for a condom and I fucked her for another thirty minutes
- (4) beneath her dress and smiled beautifully at him. She fucked him real good
- (5) You heard me. I want you to fuck me without a condom.
- (6) But if you fuck me, I’ll search the ends of the earth to find you,

Some of the verbs bear the sense of <fuck> in the sense of having sex such as ‘*she fucked him real good, I want you to fuck me without a condom*’. In these examples, the subject and the object are human nouns <N+hum><fuck.V><N+hum>. Unlike <rape.v>, where the subject is mostly man and the object is woman, both man and woman can take the two slots in <fuck.v>. In the same structure, however, the pattern can bear different sense. See example 7 and its extended context:

- (7) won’t fuck it up, I swear. // I’ll let you do it. But if you fuck me, I’ll search the ends of the earth to find you, and then you don’t want to even think about what I’ll do to you. “ Benny couldn’t believe such venom was coming from this beautiful creature. // She reached into her jacket pocket and pulled out a revolver and handed it to him. “ Here, take this, “

In (7), ‘fuck’ means to betray, although the agent and object are human. As <fuck.v> can be

an intransitive verb, the presence of object is not always necessary. Now consider the sample of <fuck.v><PREP> sequences, where an object may or may not appear:

- (8) “ I told my parents to fuck off (to leave) and got out of there.
- (9) They fuck around he’ll fuck them up (to mess) himself
- (10) And we know what you do, don’t we, Charlie? You fuck people out (to take) of money and get away with it
- (11) of school when I was 14 to go fuck around (to waste time)with my friends and live a lifestyle of crime
- (12) I got ta go fuck over ( to treat someone unfriendly) a guy who fucked me over Right-wing tendencies? I do that to fuck with people (to annoy). I thought Bush was a slimebag

In (8) ‘fuck off’ is not equipped with an object as well as ‘fuck people’ in (10). As you may see on the examples, the pattern is <fuck.v> + <PREP> as in ‘fuck off’, ‘fuck around’, and ‘fuck over’. The meaning is not always compositional (see the meaning inside the brackets). There can also be a construction where the object noun is placed after the <fuck.v> + <PREP> like ‘fuck with people’. The object noun might also take the form of MWU sequence. However, in between <fuck.V> and <PREP> there can be an object noun as in ‘fuck people out’, ‘fuck somebody up’.

- (13) people who try to kill themselves but fuck up (to make a mess) and don’t do it right?
- (14) You all want to fuck somebody up(to make a mess)? There goes a white boy; go get him
- (15) unique-to-them way to fuck this thing up(to make a mess). Even Ted Kennedy, for whom successful health care reform

### Noun

The word <fuck.n> as a noun may take the form of ‘fuck’, ‘fucker’, or ‘fuckers’. It is clear for the last two word forms (fucker and its plural form) that they have the feature of human [+HUM].

However, there are also some cases where ‘fuck’ can also refer to a human noun. See example (17) to (19):

- (16) SHEPARD # You’re a clever fuck (the boy), fireboy. And he snaps a handcuff  
 (17) Your boyfriend the dumb fuck (the boyfriend) do this? “ I shrugged;  
 (18) Jesus Christ: what kind of a sick fuck (the addressee) are you? “ Chris stepped forward

Although the form ‘fuck’ may resemble the verbal form, the part of speech in context is noun. This form highly collocates with articles ‘the’ and ‘a’. These sequences can be inserted after the verb or question words. See (20) to (25) that show the insertion of ‘the fuck’:

- (19) Vest stopped it at the shoulder, knocked the fuck out of him, though.  
 (20) I said shut the fuck up  
 (21) And one of them was getting the fuck out of the South Bay  
 (22) And you telling me again? What the fuck for? “ She slid her chair back  
 (23) WHY THE FUCK DO WE KEEP DOING WHAT WE’RE DOING?  
 (24) to American Idiot - how the fuck did I get here, man?’

Example 20 to 25 shows the insertion of ‘the fuck’ after a verb or a question word. In these examples, ‘the fuck’ is only used as a modifier. The presence of ‘the fuck’ can be deleted and the sentence can still be acceptable. This is quite different with ‘a fuck’ in (26) to (28):

- (25) The job doesn’t mean a fuck (anything) to me!  
 (26) Don’t fuck with me, Al! Don’t make a fuck (fun) out of me! You want to embarrass me  
 (27) I like seeing people lose their shit and not give a fuck (attention/care) about what other people think of them

In (26) to (28), ‘a fuck’ is not a modifier, but they play more crucial role as the core of the noun phrase in the sentence, as well as (17) to (19) previously. The use of ‘a fuck’ in these examples

functions as a replacement. Hence ‘a fuck’ cannot be deleted.

We came up to two different sequences of article plus <fuck.n>. One is obligatory and another is optional. If we refer back to (17)-(19), a word can interfere between the article and the noun as also shown by (29) to (32):

- (28) He took off his cap. “ I do n’t give a flying fuck for the Port Authority  
 (29) CHRISTIE # I do n’t give a syphallitic fuck whether you people can get along or not.  
 (30) down. The I do n’t give a rolling fuck what he does. I’ve had it  
 (31) Strapp, I do n’t give a solitary fuck if it’s from the Jurassic era

Now we come up with more sequences that can be summarized into <ART><fuck.n> and <ART><V:participle|J><fuck.n>. Symbol (|) means ‘or’. We also find that these sequences might be used as a modifier, or as a replacement of a particular item. Thus, the context will determine whether they are obligatory or optional.

### Adjective and Adverb

F-word may also take the form of present participle ‘fucking’. Here, it may take role as an adjective (to modify a noun) or adverb (to modify an adjective). Consider (33) to (38):

- (32) Maybe he shouldn’t be such a fucking, fucking idiot. # He just can’t accept it.  
 (33) he couldn’t help it. # She was so fucking beautiful. Not that he hadn’t known that  
 (34) Dad, you are one fucking weird dude. “ After explaining how serious this is  
 (35) You don’t have to cover yourself up like a fucking nun  
 (36) Close the fucking door! “ Sam would shriek back in Frank’s exact same voice  
 (37) can’t feel the fucking fork, “ his eyes wide and frightened

Here, the adjective and adverb forms of <fuck> function as modifiers for another word. The omission of ‘fucking’ in examples (33) to (38) will not generate grammatically wrong sentence. The structure is <fucking><J|N>, where J stands for adjective.



Now the question is how to employ these sequences to a corpus? We need directly exploitable formulas, and these patterns are not the answers. In COCA, you are facilitated with regular expression-like query. But how if you have your own specialized corpus? For instance, you want to know which sequence of F-words used in a corpus of movie scripts that you have collected on your own. For this, you cannot use COCA interface. You have to use off-line corpus processing software and queries that are strong enough to retrieve these sequences.

I here propose Local Grammar Graphs (LGGs) which is based on finite-state-automaton theory. Queries in LGGs (and also regular expressions) are directly exploitable in Unitex (Paumier, 2008), a multi-language platform corpus processing software. See the LGGs as shown in figure 8-10 (*see endnotes*).

Vocabularies in LGGs are put inside the box, where each box is connected from start to final state. The vocabularies can be represented as exactly as the orthographical (word) forms (such as when, who, how in figure 10), or lexical resource (lemma) codes with metasymbols, such as certain lemma or POS surrounded by angle bracket (such as 'fuck' as a verb <fuck.v>, compound noun <XN>, preposition <PRE> etc). These codes are used to explore the corpus. Further description about the codes is documented in Unitex manual by Paumier (2008).

### Challenge of Tokenization with respect to Retrieval's Precision

There are several challenges of tokenization. The first one is writing system. Some corpus processing tools are designed to process particular language(s). When it encounter texts with writing systems that are not yet encoded, it is most likely to fail to recognize the texts. The second one is the linguistic features of the language. In English, tokens are usually separated by spaces. However, in highly agglutinative or inflectional languages, this is not always the case, as different words might be concatenated (not separated by space). See example 39:

(38) The cow in the barn

In standard writing convention, each word is separated by space 'the cow in the barn', but this (39) how the orthography is visualized in highly agglutinative languages. For English texts that are written by following standard convention, corpus-processing tools can tokenize the texts smoothly. However, there are some special cases where the writing does not always follow standard writing convention (usually informal), such as texts that are obtained from movie scripts, conversation, or social media.

F-words are rarely found in formal texts. High frequency of occurrence is observed in spoken or fiction section of COCA. In these two sections, standard writing convention is not always respected. I observe that few of the surface forms of the F-words that I managed to retrieve have undergone orthographical variations.

### Spacing and Hyphen

In the COCA data that I retrieve, when F-words are used in an MWU sequence, there are at least three writing conventions. First, they are separated by space, such as 'fuck up', 'fuck you', 'fuck me'. This makes COCA identify the multi word units to be comprised of separate tokens, where each is separated by a space. See table 3:

**Table 3. Space-Separated MWU**

fuck you	fuck the
fucked up	fuck out
fuck up	fuck that
fuck it	fuck her
fuck with	fuck him

Data in table 3 clearly shows how 10 tokens <fuck> and other tokens are separated by a space. This is quite normal and the tokenization may run smooth. However, in some cases, these MWUs are tokenized as only one token as they are not separated by space. Consider the multiword expressions presented by table 4:

**Table 4. Spaceless MWU**

fuckhead	fuckball
fuckface	fuckbird
fuckups	fuckstick
fuckitall	fuckyou
fuckheads	fuckfuckfuck

Table 4 provides us with a list of 10 MWU that are not separated by space(s), such as ‘fuckoff’ (fuck off), ‘fuckhead’ (fuck head), or ‘fuckitall’ (fuck it all). Each unit is concatenated, but speakers of English are aware that these MWU are not composed by a single token. However, this is a common problem as some compound is not separated by space. The opposite might apply for ‘f u c k i n g’, where each of the characters are separated by a space (see endnotes).

The computer identifies this as seven distinct tokens, but we are quite aware that it refers to only a word. What is also important to observe is the use of hyphen(-) in separating MWUs. The concatenation of words by using hyphen can make the computer recognize the words as one token only. Consider figure 12 (see endnotes).

Just like MWUs that are not separated by space, MWUs that are separated by hyphen are also tokenized as only one token although the following MWUs ‘fuck-face’ (fuck face), ‘fuck-over’ (fuck over), or ‘fuck-you-all’ (fuck you all) are composed of more than a token.

**Unconventional Writing**

Language used in social media like twitter or facebook has carefully been studied from many perspectives such as computation, linguistics and culture. F-word is frequently used in non formal media (like social media), and one of the features that characterizes language of social media is that the writing does not always follow standard convention. Studies about unconventional variants of F-words may contribute to improve retrieval’s precision in a corpus. When this information is absent, the retrieval of all word forms of <fuck> will surely miss some unconventional variants of F-Words as presented by table 5:

**Table 5.**  
**Unconventional Writing of F-Words**

fuckin	fuuuck
fucken	fuck’em
fucka	fuckin*
fuck’n	fuckee
fuckem	fuckos

The variants as shown by table 5 are unconventional, meaning that they are not written by standard convention. This, however, is one of the features of languages used in informal situation. Consider ‘fuckin’, ‘fuckin’’, ‘fuckn’, ‘fuckin\*’, and ‘fucken’ in table 6, which are the variants of ‘fucking’. Vowel multiplication is observed in ‘fuuuck’ as the variant of ‘fuck’. The word ‘fucker’ also has several equivalences, which are ‘fucka’ for singular and ‘fuckos’ for plural. ‘fuckem’ and ‘fuck’em’ are derived from ‘fuck them’. It is interesting to observe the word ‘fuckee’ (figure 13) although it only appears twice. The word formation process adopts the derivational process of verb+ee, as in employ+ee and interview+ee. It is interesting since this form is purely the creativity of the speakers and not yet documented in standard English dictionaries.

The studies concerning hyphen, space, or unconventional writing will benefit at least in the two tasks of corpus processing; word forms identification and N-gram. One natural language processing tasks performed by corpus-processing tool is the retrieval of word forms that are derived or inflected from a lemma. How would you expect a computer to understand that, for example, ‘fuckin’ and ‘fuckn’ are the variants of ‘fucking’? The retrieval of ‘fucking’ will certainly rule out those tokens. One way is performing multiple tokens retrieval. But this means you always expect users to understand orthographical variation of a word form. Another way is to use a metasymbol in the query that will be used to retrieve orthographical variants.

Another orthographical variants that COCA cannot retrieve is the use of stars in <fuck> which are used to censor the word, like ‘f\*\*k’, ‘f\*\*\*ers’, ‘f\*\*\*ing’ because the star itself is already a metasymbol to retrieve any token. One solution is to use escape symbol. When this escape symbol is attached to star, then it retrieves the star exactly as it is written in the corpus data. I propose backslash (\). So, when you want to retrieve ‘f\*\*k’, the query will be ‘f\*\\*k’

The second task is N-Gram extraction. N-gram is useful to identify sequences of MWUs that frequently occur as a bundle. There are several areas that may benefit from the study of lexical bundles and academic writing is one of them. See

Annelie & Erman (2012), Chen & Baker (2010) or Pang (2010).

Resuming the punctuation problems, a computer will most likely to fail the extraction of sequences of MWUs that are not separated by space or separated by hyphen, as they are considered as one distinct token. Consider table 6, where the successful and unsuccessful MWUs are presented:

**Table 6.**  
**Successful and Unsuccessful Recognition of F-Word MWUs**

Successful	Unsuccessful
what the fuck	. ‘ fuck
get the fuck	do n’t fuck
who the fuck	‘s a fucking
shut the fuck	‘ ‘ fuck
where the fuck	, ‘ fuck
give a fuck	, you fucking
how the fuck	. ‘ fucking
in the fucking	do n’t fucking
want to fuck	oh , fuck
you ‘re fucking	i ‘m fucking

The result on the left column of table 7 indicates successful retrieval of three tokens for 3-Grams. The result on the left column is obtained by the same syntax query, but it also retrieves punctuation symbols like period (.), quote (‘) or double quotation (“). For the quotation (‘) it is still debatable whether a word like ‘I’m’ is composed of one or two token. In COCA, these are treated as two different tokens <I> and <’m>.

However, for double quotation, and period, either the computer misidentify these symbols as tokens, or it is the definition of token that we need to question, because (“) and (.) are identified as two distinct tokens. Information from studies concerning these issues must seriously be taken into account to improve lexico-grammatical/semantic resources used to annotate the corpus. When this information is conceived in the lexical resource, the retrieval precision is most likely to improve.

## Polarity of F-Words

In general, the semantic polarity of some expressions involving <fuck> are negative. Consider the two expressions in figure 14 (*see endnotes*).

Figure 14 presents us with two concordance lines for two different expressions. The first one is ‘go fuck yourself’, and the second one is ‘get the fuck away|out’. In these two sentences, the word ‘fuck’ does not refer to doing intercourse. However, they share similar expression ‘fuck’. Even the literal expression is equal in isolation, but in sentential context, they really differ.

In ‘go fuck yourself’, fuck is the core of serial verbs, while in ‘get the fuck away|out’, ‘the fuck’, a noun (or determiner) phrase does not function as the core of the expression. In the structure get + NP (or DP) + away, the verb ‘get’ functions as causative verb such as ‘get this car away from my lawn’. The NP, in this case ‘this car’ is an indirect object of the verb ‘get’.

In case of ‘get the fuck away’ as shown by the concordance, the NP ‘the fuck’ does not function as an indirect object. It functions as a modifier to amplify the verb. To validate this, I perform omission test for both sentence.

- (39) if you want to have sex with me... you can just go fuck yourself.
- (40) if you want to have sex with me... you can just go yourself.
- (41) Look, lady, I’ve got ta get the fuck out of here! I’ve got a date tonight
- (42) Look, lady, I’ve got ta get out of here! I’ve got a date tonight

Examples with even numbers are original sentence, and examples with odd numbers are sentences where ‘fuck’ or ‘the fuck’ is omitted. Both of the sentences are still grammatical. In (41), the sentence is still both grammatically and semantically acceptable. In (42) we might sense that the polarity is negative, but in (43) we are quite aware, when ‘the fuck’ is omitted, the polarity is no longer negative. This test has validated my proposition that the expression ‘fuck’ may function as an amplifier of the core POS. When this ‘amplifier’ is attached to modify verb, the polarity may shift to negative. Consider some verbs that are presented by figure 15 (*see endnotes*).

The verbs presented by figure 15 ('shut', 'get', 'stay', 'calm', and 'sit') are core verbs of the sentence. They are amplified by 'the fuck' to shift the polarity of the sentence. At this point, we may conclude that the amplifier may shift polarity to negative. The question that remains is whether possible that any expressions involving 'fuck' has positive polarity. In my finding, I observe that amplifiers do not always have to shift the polarity to negative. Consider the expanded context presented by figure 16 (*see endnotes*).

COCA allows the users to observe the extended context of where the target query is used. This is because the concordance is not sufficient to obtain more information that users required. The extended context shown by figure 17, the word form 'fucking' functions as an amplifier of the adjective 'good'. The adjective modifies a noun 'kush'. The noun 'kush' (not listed in standard dictionary of English) means 'marijuana'. The presence of fuck (in 'fucking good') amplifies the positive polarity of the adjective 'good'. A contradictory polarity might be observed when it amplifies different word(s). Consider figure 17 (*see endnotes*).

In the five concordance lines, 'fucking' amplifies the adjective 'idiot'. This adjective's inherent polarity is already negative, and 'fucking' amplifies it even more. Therefore, I believe that the decision to always evaluate <fuck> negatively, must be redefined seriously. One must take a closer look, at least on the linguistic context, before s/he comes into a decision. Therefore, when <fuck> is used as amplifier, the polarity will be depending on the word(s) that it modifies.

### Censorship

As it has been commented previously in the beginning of this paper, F-Words have been censored several times in the dictionary. The term 'censorship' must be understood in two different perspectives. The first one is institutional censorship and the second one is individual censorship. See Allan & Burridge (2006). In terms of institutional censorship, the censorship is performed by a particular institution (like dictionary publisher). While in individual censorship, the censorship applies to the individual itself.

The institutional censorship of, for instance, movies, is responsible to label a code to each movie with regards to its viewers segment, like G (general), PG (parental guide), R (Restricted), or A (adult). Besides labeling a movie, the censorship might also apply on the language used in the movie. The censored words are muted, or replaced by non-speech sound. The F-words are often censored by using beep along way the words are spoken. The censorship might also apply for the subtitle. Often, euphemism in the target language is used, or the F-word is not translated.

As for Individual censorship, almost all of us do it, conscious or unconsciously. The simplest representation of individual censorship is when we decide what to say and not to say. For instance, in general, it is taboo to say 'fuck'. But in certain circumstances, saying 'fuck' can be fairly acceptable. When you consult standard dictionaries of English, F-words are always marked taboo, offensive or rude. However, this does not mean that there is no rule in using <fuck>.

### Fine Tuning Boundary between negative and positive polarity

When arguing the importance of context in determining polarity, it is important to specify the term 'context' itself. The semantic interpretation of a verbal expression might be different when it is interpreted from different perspectives. Pragmatically speaking, the fine tuning boundary between negative and positive polarity of <fuck> is quite challenging to determine. Busthomi (2010) believes that dirty words can be used as intimate markers. One can curse using 'fuck you' as many as s/he likes, but when s/he is speaking to the right person, the polarity is not necessarily negative. Consider example 44:

- (43) BA1 : What's up my nigger.  
 BA2 : Hey, been long time brother.  
 Come in.  
 BA1 : You see my badass motherfucker?  
 BA2 : yeah, he right over there. Damn,  
 he doing fine in college  
 BA1 : I always know it. He is fucking  
 smart

The conversation in example 44, is between two African Americans. The use of words that

are considered taboo such as 'nigger', 'badass', 'fucking' and 'mother fucker' is obviously observed. The conversation went well, and there was no problem with the use of such 'strong' words as each might be a jargon that is acceptable in their community. However, those words are not likely in use when the participant is an outsider.

Then how do we find the use of F-words correct or not? Allan & Burrige (2006:90-112) used the term 'political correctness'. The term 'politic' does not always refer to a feud in legislation or government. Determining what to say and not to say is also politics. Therefore, the decision to use one of several existing terms is also a political decision. See (45) to (47):

- (44) The best time to have intercourse is the first two weeks after menstruation
- (45) They slept together last night.
- (46) Did you just fuck her?

Using 'to have intercourse' as in (45) might be common in medical examination report, but when having conversation with friends, using this term might seem a bit odd. Using 'fuck' might seem rude. But when this is addressed to a right person in a right situation, it is just as normal as using 'sleep together'. So, what are the variables to determine that the term we use is politically correct? In this case, it might be useful to turn to some sociolinguistics or ethnographic research. One useful concept was the classic (but still very relevant) ethnography of SPEAKING (Hymes, 1962). SPEAKING is an acronym (setting-participants-end-act of sequence-keys-instrumentalities-norms-genres) that is composed of variables that can help determine political correctness. Some general corpora, already conceive some of the variables. For instance, BNC conceives the information about the speakers' age, sex and dialect. COCA has already included genre information. However, these corpora are general corpora. To help setting the fine tuning boundary of positive and negative polarity, either the corpora's features are improved, or new specialized corpora are created.

## CONCLUSION

The distribution of F-words across corpora has shown that the frequency is highest in COCA. Careful examination of F-words concordances in

COCA has provided us with the information about the MWU sequences. It shows the flexibility of F-words in terms of POS. The sequences of MWUs are shown by three LGGs in figure 9 to 11, which are directly exploitable in Unitex. The three LGGs can be summarized in one LGGs. When users want to retrieve <fuck> as noun, verb, adjective, or adverb, used in a sequence, only one retrieval is required. See figure 18 (*see endnotes*).

This research also suggests that different senses that may be derived of F-words used in MWUs are the outcome of its collocation. The result of the retrievals also shows that the use of F-words is optional in some case because they are used as amplifier. In this case, the polarity can be positive or negative depending on a word that it specifies. By the end of the paper, I also show that the positive or negative polarity of F-words might be determined from different variables. The lexico-grammatical and componential features of F-Words might be useful in determining the semantic polarity. However, to decide the political correctness of using F-words on discourse level, different variables are required such as one from Hymes' SPEAKING (1962).

## REFERENCES

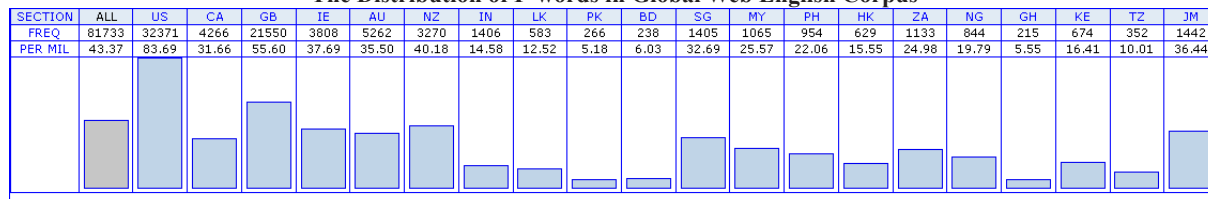
- Allan, K., & Burrige, K. (2006). *Forbidden Words: Taboo and The Censoring of Language*. Cambridge: Cambridge University Press.
- Annelie, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes* 31.2, 81-92.
- Brewster, S. (2013). Saying the 'F word... in the nicest possible way': augmentative communication and discourses of disability. *Disability & Society*, 28(1), 125-128.
- Busthomi, Y. (2010). On Dirty Words: An Attempt to Intimate The Conundrum Of The Sin. *Bahasa dan Seni*, 1-9.
- Chen, Y.-H., & Baker, P. (2010). Lexical Bundles In L1 And L2 Academic Writing. *Language Learning & Technology* (2), 30-49.
- Chiorean, S. (2014). *The Case of Fuck in Corpus of Contemporary American English (Thesis)*. Stockholm: Stockholm University Press.
- Davies, M. (2008). *American Corpus*. Retrieved from The corpus of contemporary American English

- (COCA): 385 million words, 1990-present: <http://www.americancorpus.org>
- Davies, M. (2010). The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Language and Literary Computing*, 447-464.
- Fairman, C.-M. (2007). Fuck. *Cardozo Law Review*, 28(4), 1711-1772.
- Firth, J. (1957). *Papers in linguistics 1934-1951*. London: Oxford University Press.
- Harding, N., Ford, J., & Fotaki, M. (2013). Is the 'F'-word still dirty? A past, present and future of/for feminist and gender studies in Organization. *Organization*, 20(1), 51-65.
- Hobbs, P. (2012). Fuck as a metaphor for male sexual aggression. *Gender and Language*, 7(2), 149-176.
- Howe, R. (2012). *The Use of Fuck: A Sociolinguistic Approach to the Usage of Fuck in the BNC and Blog Authorship Corpus*. Michigan: Michigan University Press.
- Hughes, G. (2006). *Encyclopedia of Swearing*. New York: M.E Sharpe Inc.
- Hymes, D.-H. (1962). The ethnography of Speaking. In T. G. (eds), *Anthropology and Human Behaviour*. Washington, D. C: Anthropology Society of Washington.
- Jaworska, S., & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990-2009. *Discourse & Society* 23(4), 401-431.
- Johnson, S. (2003). *Samuel Johnson's Dictionary*. London: Walker & Company.
- Kravis, N. (2013). Fuck Redux A Review And Commentary. *Journal of the American Psychoanalytic Association* 61(3)527-534.
- Leigh, M., & Lepine, M. (2005). *Advanced Swearing Handbook*. Sussex: Sumersdale Publisher.
- Ludeling, A., & Kyto, M. (2009). *Corpus Linguistics: An International Handbook*. Berlin and New York: Walter de Gruyter.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, A., & Xiao, Z. (2004). Swearing in modern British English: the case of fuck in the BNC. *Language and Literature* 13(3), 235-268.
- McEnery, A., Baker, J.-P., & Hardie, A. (2000). Assessing claims about language use with corpus data: Swearing and abuse. *Language and Computers*, 30, 45-56.
- McEnery, T. (2006). *Swearing in English*. New York: Routledge.
- Nadar, F., Wijana, I., Poedjosoedarmo, S., & Djawanai, S. (2005). Penolakan dalam Bahasa Inggris dan Bahasa Indonesia. *Humaniora* (17), 166-178.
- Norricks, N.-R. (2012). Swearing in literary prose fiction and conversational narrative. *Narrative Inquiry*, 22(1), 24-49.
- O'Keefe, A., & McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge.
- Olive, R., & Thorpe, H. (2011). Negotiating the 'F-word' in the field: Doing feminist ethnography in action sport cultures. *Sociology of Sport Journal* 28(4), 421-440.
- Pang, W. (2010). Lexical Bundles and the Construction of an Academic Voice: A Pedagogical Perspective. *Asian EFL Journal*, Vol 47.
- Paumier, S. (2008). *Unitex Manual*. Paris: Universite Paris Est Marne La Valee & LADL.
- Sheidlower, J. (2009). *The F Word*. Oxford: Oxford University Press.
- Sparrow, J. (1962). Regina V Penguin-Books-Ltd-An Undisclosed Element In The Case. *Encounter* 18(2), 35-43.

#### (Endnotes)

- 1 like *fuck*, *fucking*, *fucker* which are well understood as 'dirty word'. It is often considered a taboo subject to discuss in everyday communication. However, the works of Busthomi (2010), Howe (2012) Brewster (2013) and Chiorean (2014) indicate that this topic is academically worth to discuss from a wide range of topics.
- 2 Speech corpus, transcription of refusals (Nadar, 2005) published by Humaniora Journal Gadjah Mada University
- 3 Despite its title 'American English' this website provides not only American English Corpus, but also other world Englishes, such as British and Australian English, and also where English is used as foreign language like Hongkong (China), the Phillipine, Singapore and etc.
- 4 Syntax here refers to strings of characters to run commands on computer, not the study of sentence structure in linguistics. However, in corpus processing, this syntax is a crucial importance to retrieve certain linguistic forms as queried by the corpus investigator
- 5 [www.americancorpus.org](http://www.americancorpus.org)
- 6 <http://corpus.byu.edu/coca/>
- 7 Cues in the form of words and clauses are referred as Multi Word Units (MWUs)

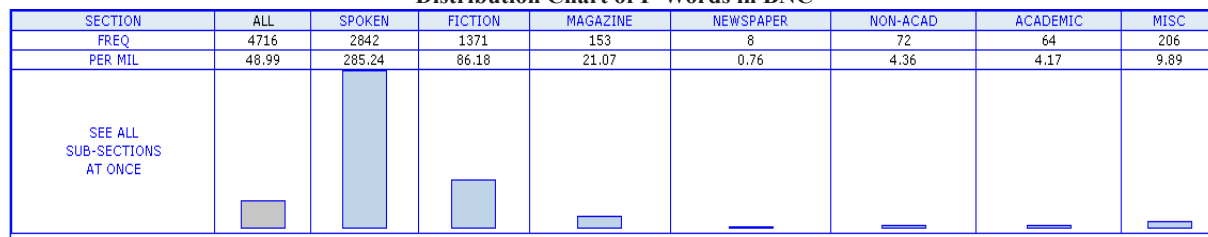
**Figure 1.**  
**The Distribution of F-words in Global Web English Corpus**



**Figure 2.**  
**The Retrieval of F-Words in BNC**

	<input type="checkbox"/>	CONTEXT	FREQ
1	<input type="checkbox"/>	FUCKING	3039
2	<input type="checkbox"/>	FUCK	1301
3	<input type="checkbox"/>	FUCKED	162
4	<input type="checkbox"/>	FUCKIN	51
5	<input type="checkbox"/>	FUCKER	48
6	<input type="checkbox"/>	FUCKS	23
7	<input type="checkbox"/>	FUCKERS	17
8	<input type="checkbox"/>	FUCK-ALL	12
9	<input type="checkbox"/>	FUCKED.	10
10	<input type="checkbox"/>	FUCKED-UP	5

**Figure 3.**  
**Distribution Chart of F-Words in BNC**



**Figure 4.**  
**The Retrieval of F-Words in Canadian English Corpus**

	<input type="checkbox"/>	CONTEXT	FREQ
1	<input type="checkbox"/>	FUCKING	127
2	<input type="checkbox"/>	FUCK	112
3	<input type="checkbox"/>	FUCKIN	36
4	<input type="checkbox"/>	FUCKED	20
5	<input type="checkbox"/>	FUCKS	4
6	<input type="checkbox"/>	FUCK-YOU	4
7	<input type="checkbox"/>	FUCKER	4
8	<input type="checkbox"/>	FUCKEL	4
9	<input type="checkbox"/>	FUCKERS	2
10	<input type="checkbox"/>	FUCKED.	2

Figure 5.  
Distribution Chart of F-Words in Canadian English Corpus

SECTION	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-FICTION	ACADEMIC	MISC
FREQ	14	182	65	3	13	49	0
PER MIL	2.45	48.83	6.52	0.23	5.18	3.39	0.00
SEE ALL SUB-SECTIONS AT ONCE							

Figure 6.  
The Retrieval of F-Words in COCA

1	<input type="checkbox"/>	FUCK	7936
2	<input type="checkbox"/>	FUCKING	7483
3	<input type="checkbox"/>	FUCKIN	2102
4	<input type="checkbox"/>	FUCKED	1366
5	<input type="checkbox"/>	FUCKER	376
6	<input type="checkbox"/>	FUCKS	235
7	<input type="checkbox"/>	FUCKERS	146
8	<input type="checkbox"/>	FUCKED-UP	131
9	<input type="checkbox"/>	FUCK-UP	46
10	<input type="checkbox"/>	FUCKEN	44

Figure 7.  
Full Concordance of 'fuck'

1	2012	FIC	Bk:Grown-upKindPretty	A	B	C	to hiss, " Liza? Liza-Little? Mosey is here. Shut the fucking <b>fuck</b> up. " # She kept yelling through my hand. I could feel the
2	2012	FIC	Bk:RobertLudlumsThe	A	B	C	speed of liquid flame. # Janson kicked it from his hand. # " <b>Fuck!</b> " # Doug Case rubbed his wrist where Janson's boot had connected.
3	2012	FIC	Bk:WildThingNovel	A	B	C	or just too polite, to admit they're not numb. # Well, <b>fuck</b> it. Save the lido for someone too scared to lie. # I twist
4	2012	FIC	Bk:WildThingNovel	A	B	C	as I know the highest part of the ship you can actually stand on. <b>Fuck</b> knows what it has to do with fire. # The sun's setting,
5	2012	FIC	Bk:WildThingNovel	A	B	C	walk up to her. # " I'm Lionel Azimuth. " # " <b>Fuck</b> off. " # I don't take it personally. I look like a
6	2012	FIC	Bk:WildThingNovel	A	B	C	nine a.m. -- late to get a line down, like Charlie Brisson gives a <b>fuck</b> . He's not out on this bullshit lake in the middle of the fucking
7	2012	FIC	Bk:WildThingNovel	A	B	C	that's bullshit, and Good Lisa never existed in the first place, but <b>fuck</b> -- he just misses her so much. # The sobs break out of him
8	2012	FIC	Bk:WildThingNovel	A	B	C	. It's as empty as Brisson's motherfucking life. # Heh-heh-heh. # <b>Fuck</b> crappie. Fuck fucking walleye. After Brisson found out Lisa was fucking Robin,
9	2012	FIC	Bk:WildThingNovel	A	B	C	as empty as Brisson's motherfucking life. # Heh-heh-heh. # Fuck crappie. <b>Fuck</b> fucking walleye. After Brisson found out Lisa was fucking Robin, Lisa swore to
10	2012	FIC	Bk:WildThingNovel	A	B	C	, thank you, Jesus. Yanks it loose to row for shore, where <b>fuck</b> it -- fuck all of it -- he can start drinking again. # Back

Figure 8.  
LGGs for <fuck> as a Verb

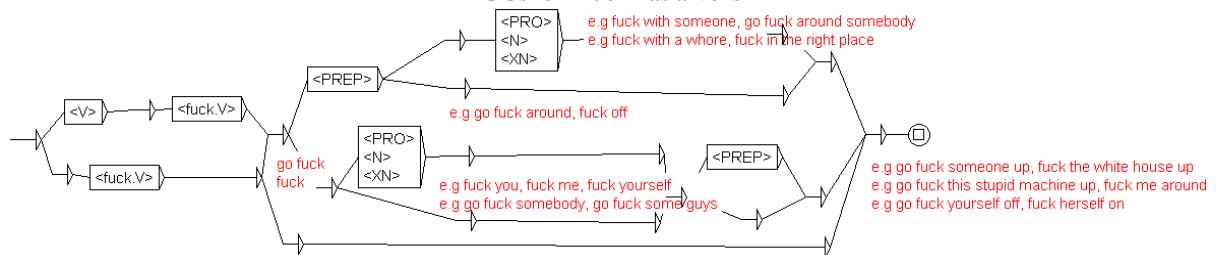




Figure 9.  
LGGs for <fuck> as a Noun

e.g where the fuck, how the fuck  
 e.g get the fuck up, shut the fuck down  
 e.g sell it the fuck up, give the white house a fuck about  
 e.g give a fuck about, get a fuck with  
 e.g give a flying fuck, got a good fuck up

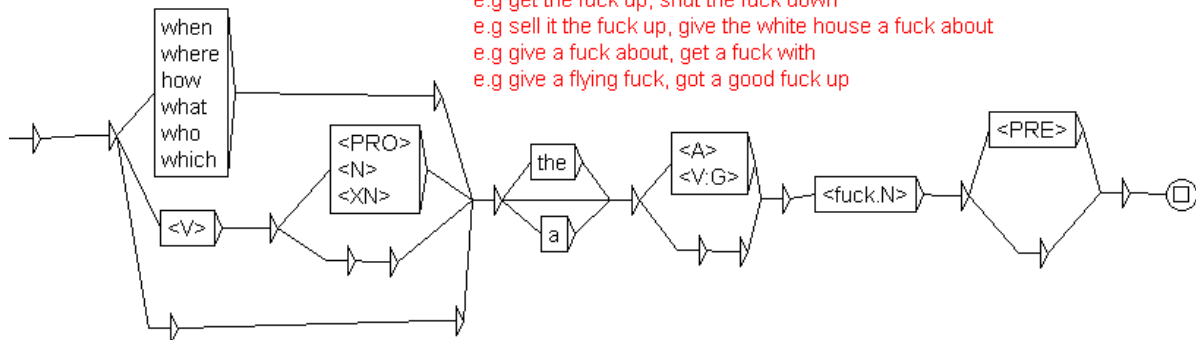


Figure 10.  
LGGs for <fuck> as Adjective and Adverb

e.g fucking idiot, fucking beautiful, fucking lovely  
 e.g fucking door, fucking business, fucking lovesong

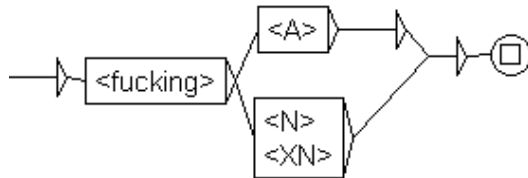


Figure 11.

<fuck> where each characters are separated by space

1	2010	ACAD	AmerScholar	A	B	C	last possible day, who aren't likely to rearrange the keyboard to spell <b>F U C K</b> THIS or to accidentally " lock " themselves in the
2	2005	FIC	Mov:Capote	A	B	C	d ing and there's no <b>f u c k</b> ing end in sight.... No

Figure 12. Hyphenated MWU

1	2012	FIC	Bk:BloodBullets	A	B	C	. In my world, nights that started strange usually crash-landed in the town of <b>Fucked-Upville</b> before they were said and done. # The stereo kicked in with the winsome
2	2012	FIC	SouthernRev	A	B	C	to associate yourself with me, such was the power, the dominance of my <b>fucked-upedness</b> ... Another theory had to do with television. Perhaps we'd all watched too
3	2012	ACAD	AmerScholar	A	B	C	the Forverts. But since then Max has turned into the angry Jew, the <b>fuck-you-both</b> Jew, not an un-Jew but an anti-Jew, not abashedly assimilated but proudly so
4	2011	FIC	Bk:DubiousSalvation	A	B	C	words is that my father had a filthy mouth. His favorite words were " <b>fuck-face</b> " and " cunt. " Prolonged exposure to my father's foul language led
5	2009	MAG	RollingStone	A	B	C	was always a classical-music guy. I was never a peacemaker; I was a <b>fuck-you</b> rightwinger. I've got a weird view of American history that I think is
6	2008	FIC	Bk:GivenDay	A	B	C	, and, of course, the Gusties, the city's most powerful and <b>fuck-out-of-their-minds</b> street gang, who headquartered in South Boston but spread their tentacles throughout t
7	2008	FIC	Bk:TenthCase	A	B	C	forever earning himself the aka Herman Greencard; a bankruptcy specialist known in-house as " <b>Fuck-the-Creditors</b> " Feinblatt; an older guy who did nothing but chain-sm
8	2008	FIC	LiteraryRev	A	B	C	seductive movements with their fannies. At the beach we're all equal, us <b>fucked-overs</b> and them. We're better, even, because we don't have the
9	2007	FIC	Bk:MichaelTolliver	A	B	C	attendance at Queer as Folk night. Not to mention the traffic snarls and the <b>fuck-you-all</b> matre d's and the small-town queers who brought their small-town fears to the Cast
10	2000	FIC	Iris	A	B	C	beyond our fences, # rows of mendicant brownstones # and metal-latticed shops beneath a <b>fuck-scribbled</b> # platform from which we felt the acrid shiver, # the hellish clan

Figure 13.  
Fuckee

CLICK FOR MORE CONTEXT		<input type="checkbox"/>	[?]	SAVE LIST	CHOOSE LIST	-----	CREATE NEW LIST	<input type="text"/>	[?]
1	2000	FIC	NewEnglandRev	A	B	C	got to decide, Danger Man. Are you the fucker or are you the <b>fuckee</b> ? These girls, they don't got a choice. They're nothing but		
2	1991	FIC	Mov:JFK	A	B	C	's fuckin' us now! And that fuckin' zealot Bobby Kennedy is the <b>fuckee</b> ! The nerve of that little asswipe closing the camps. Took all our C-4		

Figure 14.  
Negative Semantic Polarity of F-Words

1	2011	FIC	Bk:BreakingRulesNovel	A	B	C	visit? It's time you moved into big-boy housing. " # " <b>Go fuck yourself</b> , " Dan said, genuinely pissed. Apparently Izzy had trod on a
2	2011	FIC	Bk:BreakingRulesNovel	A	B	C	on? # Izzy: (throwing a punch) Why don't you <b>go fuck yourself</b> ...? # " Y'okay? " Izzy asked Jenk instead as the other
3	2011	FIC	Bk:BlackLightNovel	A	B	C	everything? " # I roll my eyes. " Money's great. <b>Go fuck yourself</b> . " # He laughs. " Look, do yourself a favor,
4	2011	FIC	IowaRev	A	B	C	the other girls, a Swede, says " No, " meaning, <b>go fuck yourself</b> acne face. The brassy-haired girl holds her fingers to the Swede's lips
5	2010	FIC	Bk:HubMatter	A	B	C	if you want to have sex with me... you can just <b>go fuck yourself</b> . " " These are acceptable results, " Vekredi told Nashira as he
1	2012	FIC	Bk:YouCameBackNovel	A	B	C	, " he said. " Let's apply for Canadian citizenship. <b>Get the fuck</b> out of Jesusland. " # " New Zealand, " she said. "
2	2012	FIC	SouthwestRev	A	B	C	, Ben would whisper to me, " I can chant too: <b>Get the fuck</b> away from me, get the fuck away from me, get the fuck away
3	2012	FIC	SouthwestRev	A	B	C	" I can chant too: Get the fuck away from me, <b>get the fuck</b> away from me, get the fuck away from me, " while -
4	2012	FIC	SouthwestRev	A	B	C	the fuck away from me, get the fuck away from me, <b>get the fuck</b> away from me, " while - not wanting to cause offense - he
5	2011	FIC	FantasySciFi	A	B	C	sky behind him. " Look, lady, I've got to <b>get the fuck</b> out of here! I've got a date tonight! " " You think

**Figure 15.**  
**The Retrieval of Verb + the fuck + Any Token**

1	<input type="checkbox"/>	SHUT THE FUCK UP	190
2	<input type="checkbox"/>	GET THE FUCK OUT	123
3	<input type="checkbox"/>	GET THE FUCK OUTTA	28
4	<input type="checkbox"/>	GET THE FUCK AWAY	23
5	<input type="checkbox"/>	GET THE FUCK OFF	16
6	<input type="checkbox"/>	GET THE FUCK UP	9
7	<input type="checkbox"/>	STAY THE FUCK AWAY	8
8	<input type="checkbox"/>	CALM THE FUCK DOWN	7
9	<input type="checkbox"/>	GETTING THE FUCK OUT	7
10	<input type="checkbox"/>	SIT THE FUCK DOWN	6

**Figure 16.**  
**Expanded context of ‘fucking good’**

Publication information	Boston : Mariner Books,
Title	Mule :a novel of moving weight
Author	D'Souza, Tony.

**Expanded context:**

# When I came back the three and a half hours from the Sacramento airport with Mason in the car, still reeling in my head from the busy city, Kate had a neat blunt rolled for him and they got stoned on the porch right away. All the two of them did, as they smoked under our tall trees, was laugh like a pair of old hyenas. When I asked them what they were laughing about, Mason grinned at me and said, " Poor James doesn't even know. It's because this kush is so fucking good." # " Look at you now," he said to me, " a mountain man and all. In a motherfucking flannel jacket just like a goddamn lumberjack." # I crossed my arms where I stood in the yard and beamed at him. # " How much do you pay for an eighth of this?" Mason asked Kate. Kate batted her eyes at him and said, " Man, don't you know I grew up out here? I get this shit

**Figure 17.**  
**Full Concordance of ‘fucking idiot’**

1	2012	FIC	Bk:WildThingNovel	A	B	C	out to the woods. Maybe he shouldn't be such a fucking, <b>fucking idiot</b> . # He just can't accept it. It's like some new Lisa
2	2011	FIC	AmerTheatre	A	B	C	know what he means. He doesn't mean us. # ANITA: <b>Fucking idiot</b> . (Pause) ralc, I think Maini stays because she has no other
3	2011	FIC	AmerTheatre	A	B	C	.(Beat) I just made no sense. I sounded like a <b>fucking idiot</b> just now. Moron! Malte comes down, looking all... she comes
4	2011	FIC	Antipodes	A	B	C	Barking. # His voice, high-pitched with drink and fear, swearing. <b>Fucking idiot</b> , he's saying, Fucking idiot. # He is tripping away down the
5	2011	FIC	Antipodes	A	B	C	with drink and fear, swearing. Fucking idiot, he's saying, <b>Fucking idiot</b> . # He is tripping away down the hall. # You stand right where

**Figure 18.**  
**LGGs for F-Words**

