Diachronic Analysis of the Profane Words in English Song Lyrics: A Computational Linguistics Perspective

Flora Goyak[1], *Mazura Mastura Muhammad[2], Muhammad Fadzllah Zaini[3]
Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris
35900 Tanjung Malim, Perak Darul Ridzuan, Malaysia
e-mail: mflora89@yahoo.com[1], mazura@fbk.upsi.edu.my[2], mfadzllah@fbk.upsi.edu.my[3]

Wesam Mohamed Abdelkhalek Ibrahim
Department of Basic Sciences, Community College,
Princess Nourah bint Abdulrahman University, Saudi Arabia
Department of Foreign Languages
Faculty of Education, Tanta University, Egypt
e-mail: wmibrahim@pnu.edu.sa

Ardi Gunsuh
Talent Corporation Malaysia Berhad
e-mail: ardi.gunsuh@talentcorp.com.my

*Corresponding author: mazura@fbk.upsi.edu.my

Abstract

This diachronic study aims to explore the linguistic phenomena of the verb *f-word* in English song lyrics across genres and time via various corpus computational tools. A specialised corpus named Diachronic Corpus of English Song Lyrics (DCOESL) consisting of Country, Pop, Rhythm and Blues (R&B), and Rock genres from the years 1960 to 2009, was built for the analysis. Linguistics analysis of English song lyrics corpus was used as the research design. Computational corpus instruments were adopted to generate data. The findings reveal that corpus computational tools it has provided an avenue for researchers to explore languages across time. Additionally, the study shows that *f-word* in English song lyrics experience ascending trend since the 1980s, with highest occurrences in R&B (38pmw). *F-word* in DCOESL has strong collocational strength with personal pronoun me (17321pmw), MI=3.442. Personal pronoun me is very significant to the node *f-word*, T-score=3.274. *F-word* in DCOESL has highest significant lexical association with *f-word* in the spoken register of

COCA, G²COCASPOKEN=102.40, df=1, p<.0001. It exhibits that the highest occurrences of *f-word* in DCOESL reflects social actions and a high preference for simple present tense, and simple sentence structure. In conclusion, the computational corpus analysis of *f-word* in English song lyrics has found that *f-word* prominently co-occur with personal pronoun in simple sentence structure and in simple present tense, in order to mirror English conversational discourse. The implication of this study is English song lyrics, especially from R&B genre, are a potentially authentic corpus resource for exploring spoken English.

Keywords: English song lyrics, corpus computational tools, diachronic study,
f-word, spoken English

Introduction

Computational Linguistics is an interdisciplinary field that provides a computational perspective of the natural language and it automates various linguistics tasks which previously were carried out manually such as text analysis, lexical mapping and information extraction and retrieval. Researchers are now able to explore and analyse a massive amount of natural occurring language such as song lyrics.

Songs are an effective teaching tool through which learners can learn language in an entertaining, active and stimulating way. The combination of language and entertainment shows to learners that language learning can be fun (Domoney & Harris, 1993; Lo & Fai Li, 1998). Besides that, the repetitive nature of songs promotes memory retention since the rhythm and rhyming patterns help students to recall lexis and syntactical chunks (Schoepp, 2001). Additionally, songs enhance the use of English language that learners will face in real life contexts and students are able to hear vocabulary and phrases in a natural and meaningful context (Eken, 1996). Scholars such as Bartle (1962), Richards (1969) and Jolly (1975) have been asserting for use of songs as an affective language learning tool for both their linguistics advantages and motivational impacts that they are capable to generate in language learners. This is aligned with the second language acquisition theory namely the Affective Filter Hypothesis developed by Krashen (1982) as a foundation that relates applied linguistics study as a contribution to teaching and learning practice of the English language.

The use of taboo words is a fraction of daily language use. Recent studies by Goddard (2014, p. 55), Ljung (2011), McEnery (2006), and McEnery and Xiao (2003) have addressed the taboo *f-word* and various aspects of *f*-word. However, the aforementioned studies focus on written and spoken registers that do not include song lyrics. For instance, McEnery and Xiao (2003) investigated the use of *f-word* in written section of British National Corpus (BNC) with respect to a subset of the metadata namely domain, author gender, author age, audience gender, audience age, audience level, reception status, medium of text and date of creation (p. 504). The researchers discovered four main characteristics of written *f*-word; it is used among lower level of audience comparative with speech from the lower class, it is a marker of male readership or authorship comparative to a marker of male speaker, used more frequently by younger writers comparative to as a word regularly spoken by younger speakers, and correlate with more informal types of writing particularly imaginative writing (McEnery & Xiao, 2003, p. 511). Other than that, McEnery (2006) used the

Lancaster Corpus of Abuse (LCA) which is based on the spoken BNC to investigate categories of Bad Language Word (BLW) use and gender. They discovered that male speakers use *f-word* twice more frequent (68.28pmw) than the females (32.75pmw) with LL value of 48.98 (p. 29). The aforementioned intriguing findings are unlimited to linguistics uses of *f-word,* by including the elements of social class, gender, and age to their investigation. The current study developed a corpus song lyrics namely the Diachronic Corpus of English Song Lyrics (DCOESL) in massive size to cater for the available slit in including song lyrics for the linguistics investigation of *f-word*. DCOESL also opens opportunities for exploring the origin of *f-word* down the history and ways native speakers use the taboo word.

Previous corpus studies on song lyrics by Logan et al. (2004) and Taina (2014) include words distributions by genre. Logan et al. (2004) included Country, Reggae, Newage, Rap and Rock, whereby swear words namely *nigga* and *shit* are identified in Rap (p. 2). Although Logan et al. (2004) claimed that their study is mainly on semantic analysis, their results and discussions are more towards clusters classification of popular songs through lyrics. A more detailed discussion on linguistics items such as collocations and sentence patterns would be intriguing. For instance, although lyric-specific words such as *I'm* and *love* occur in each genre, detailed discussion on describing the semantic reasons behind these occurrences are not provided in their research. The current study employs semantic analysis to identify and discuss in detail the semantic categories and meanings of *f-word* in song lyrics. Taina (2014) discovered that *f-word* and shit are common in Thrash Metal, and thus assumed to be the stylistic feature characteristic of the subgenre (pp. 49-54). Note that the study conducted by Taina (2014) is based on metal genre. The disadvantage of a single genre corpus is findings are not able to be wholly generalised to linguistics features in general. While the findings can be representative of linguistics features of specifically metal songs in general, the corpus is not reliable enough to represent similar linguistics features of other music genres or registers in the English language. As asserted by Lindquist (2009), representativity is comparable because it differs among corpora and certainly not absolute. The usefulness of the findings can be submitted for comparative analysis with general reference corpora. For example, the current study utilised BNC for comparative analysis. The grammatical aspects namely collocations and sentence patterns, and statistical test results are used to generalise song lyrics as a spoken-like register of the English language. These studies do include swear or profanity words, but lack of in depth focus on each lexical item. Bridle (2018), Motschenbacher (2016), Saarinen (2013), and Petrie et al. (2008) are among other studies that is based on a single genre corpus. These researchers are focusing on Blues, Pop, Metal, and Rock respectively. For this study, DCOESL was analysed and the focus is on the phenomena of *f-word* across four popular music genres and time. The use of corpus linguistics in this research benefits teachers who want to use songs as effective authentic teaching and learning materials for their students; more specifically songs to avoid in their teaching.

Methodology

Corpus Description: Diachronic Corpus of English Song Lyrics (DCOESL)

Diachronic study of language via corpus-driven approach was employed to analyse *f-word* that are used in English song lyrics across genres and time. Diachronic study was selected to diachronically study the emerging linguistic changes of most frequent pronouns in English song lyrics across the span of five decades and four music genres. Corpus-driven approach was employed to analyse the large and principled collection of naturally occurring English song lyrics that function as the empirical basis in studying linguistic phenomena of pronouns in the diachronic corpus. Tognini-Bonelli (2001) described that with this type of approach, the "descriptions aim to be wide-ranging with respect to corpus evidence" and linguistic categories are derived systematically from the "recurrent patterns and the frequency distributions that emerge from language in context" (p. 87). In this study, *f-word* is under lexical category. The *f-word* was investigated for frequency distribution and collocational analyses. Data interpretation and discussion were carried out in respect to the emerging findings from the diachronic corpus.

Diachronic Corpus of English Song Lyrics or DCOESL was built by compiling and organizing a total of 5000 song lyrics from the years 1960 to 2009 comprising of 25 songs for every year of each genre, generating of approximately one million running words. In order for texts to be considered as a corpus, there is no minimum or maximum size required (Sinclair, 2015). DCOESL is intended to be a dense song lyrics corpus for the purpose of covering a wider range of emerging linguistics variations than what smaller corpora could do. Accordingly, DCOESL must be large enough to sufficiently represent the occurrence of *f-word* in English song lyrics.

Specialised corpus type is selected for this study. A specialised corpus is inclined to be domain or genre specific (McEnery et al., 2006). DCOESL consists of four distinct music genres namely Country, Pop, Rhythm and Blues (R&B), and Rock. The song lyrics from the aforementioned genres were selected because of their popularity and free access via the internet. The songs for this study were carefully chosen based on popularity from every decade and genre. Among the many options available in the cyberspace, chart songs rank highly in popularity (North et al., 2004). Hence, the Top 100 Billboard Charts (www.billboard.com), an online extension of the Billboard magazine, serves as a suitable platform for retrieving popular song lists. Billboard is a well-known magazine for publishing pieces such as news, opinion, reviews, styles, videos, and music charts. Song lyrics needed for this study were compiled from MetroLyrics (www.metrolyrics.com), a licensed lyric-based website that provides database of over one million songs performed by over sixteen thousand artists. Table 1 contains numerical description of DCOESL.

Table 1

Numerical Description of DCOESL

Genres	Year	Number of song lyrics for every year	Tokens
Country	1960-2009	25	290 278
Pop		25	357 770
Rhythm and Blues (R&B)		25	460 545
Rock		25	303 828
Total	50	5 000	1 412 601

Research Instruments

McEnery et al. (2006) asserted that computer helps to process and manipulate corpus data rapidly at minimal cost, avoids human bias to achieve reliability, and allows further automatic processing to be performed for various metadata enrichment. Computer-aided analysis via the aforementioned five corpus analysis instruments, makes it possible for this research to be carried out. First, AntConc is a freeware corpus analysis toolkit for concordance and text analysis developed by Professor Laurence Anthony. The software comprises of seven corpus tools namely concordance, concordance plot, file view, clusters or n-grams, collocates, word list, and keyword list. In this study, AntConc was utilised to generate frequency lists of *f-word* and its adjacent collocations, and concordance lines. Second, CLAWS POS Web Tagger is a reliable program for automated part-of-speech tagging that was developed by University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University. Song lyrics in DCOESL were tagged by using C7 Tagset, including *f-word*, which is tagged as VV0, a tag that stands for base form of lexical verb. Third, Lancaster Statistics Tools online was used to automatically calculate the complex formulae of three statistical tests of significance namely Log-likelihood (G^2), Mutual Information (MI), and T-score. The Chi-square (X2) test was conducted via Social Science Statistics online calculator. Last, similar to CLAWS POS Web Tagger, USAS was developed along with its own tagset. USAS consists of 21 major discourse fields. In the current study, USAS functioned to identify semantic categories of *f-word* and its most frequent adjacent collocates in DCOESL. Prominent discourse fields tagged in DCOESL are S (social actions, states, and processes), B (the body and the individual), and E (emotion).

Reliability and Validity

To ensure that the occurrences of verb *f-word* in DCOESL are not mainly based on raw frequencies alone, four statistical measurements were used namely normalization of frequency counts, Log-likelihood (G^2), Mutual Information (MI) test, and T-score test.

First, the raw frequencies of DCOESL were normalised to 1 million words or per million words (pmw). Second, comparisons of frequency counts between DCOESL and the reference corpora BNC and COCA were conducted to determine the extent the findings can be generalised to the language as a whole. G^2 was carried out to measure of strength of associations between *f-word* in DCOESL and *f-word* in the reference corpora. Third, to measure collocational strength, Mutual Information (MI) test was carried out. The higher the MI score, the stronger the link between two items (McEnery et al., 2006, p. 56). Finally, T-score test was done to measure the significant of a collocate to the node. T-score test in this study is used to identify words that are vital for the node *f-word*.

Findings and Discussion

Distribution of *F-word* and *F-word* Variations in DCOESL

In this section, both quantitative and qualitative results of DCOESL are presented and discussed. The findings consist of three types namely frequency count, annotation, and comparative analysis. First, the frequency counts of *f-word* and its adjacent collocates are tabulated in the forms of tables and graphs. Next, song lyrics in DCOESL underwent part-of-speech and semantic tagging for annotations. Last, comparative analysis of frequency counts with general corpora namely the British National Corpus (BNC) and Corpus of Contemporary American English (COCA).

With the aim to begin the analysis, the number of counts for *f-word* in DCOESL were generated via AntConc. Table 2 shows the frequency and percentage distributions, and Table 3 shows the frequencies of *f-word* variations in DCOESL. Figure 1 illustrates the distribution of *f-word* across genres and time in DCOESL.

Table 2

Frequencies and Percentage Distributions of F-word in DCOESL

Genres	RF	NF	%
Country	0	0	0
Pop	15	11	0.0011
R&B	55	39	0.0039
Rock	34	24	0.0024
Total	104	74	0.0074

Note. RF = Raw Frequency, NF = Normalised Frequency, % = Percentage of Distribution. The figures for % are rounded off to the nearest four decimal places.

The *f-word* was searched separately for its variants which are *fuck, fucked, fucks, fuckin(g),* and *fucker(s)* in order to attain clearer view of their distributions in each music genre over a span of five decades of the whole DCOESL.

Table 3

Distribution of F-word Variants in DCOESL

Form	POS Tag	DCOESL				vs. BNC		vs. COCA	
		Genre	RF	NF	G^2	Sig. Level	G^2	Sig. Level
fuck	VV0	County	0	0	79.55	<0.0001	62.71	<0.0001
		Pop	15	11				
		R&B	55	39				
		Rock	34	24				
fucked	VVD	County	0	0	10.97	<0.001	10.97	<0.001
		Pop	4	3				
		R&B	10	7				
		Rock	3	2				
fucks	VVZ	County	0	0	0.25	<1.0	0.36	<1.0
		Pop	0	0				
		R&B	0	0				
		Rock	0	0				
fucking	VVG	County	0	0	0.95	<1.0	1.06	<1.0
		Pop	1	1				
		R&B	0	0				
		Rock	1	1				
fucker(s)	NN1, NN2	County	0	0	7.30	<0.01	5.85	<0.05
		Pop	2	1				
		R&B	2	1				
		Rock	9	6				

As can be seen from the table, for all the word forms under investigation, the lexical verb *f-word* has the highest occurrences with a total of 74 counts per million words (pmw). The difference between DCOESL against BNC and COCA is statistically significant at the level <0.001. *F-word* in DCOESL occurs about 24 times more frequently than in BNC, and about 10 times more frequently than in COCA. The second most contrast is found for the past tense form, f**ked, which is used about 12 times more frequent than f**ked in BNC and COCA.
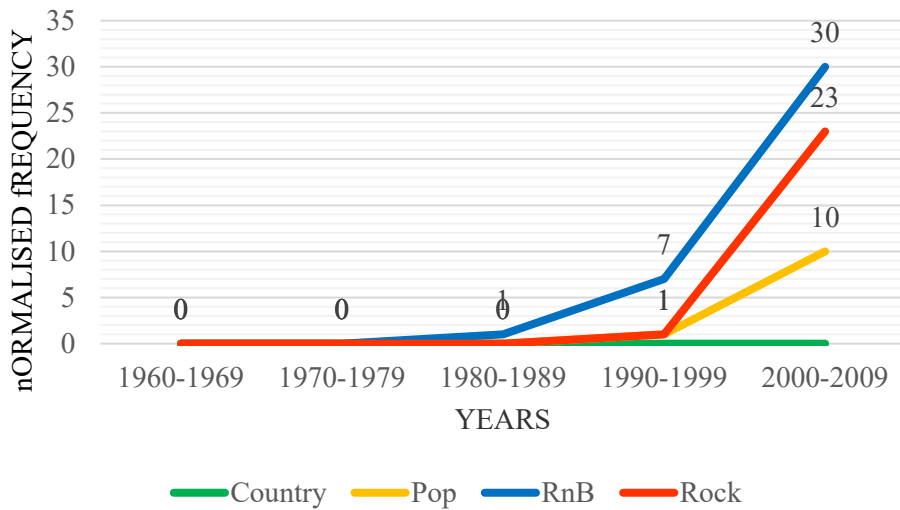
Figure 1. Distribution of Lexical Verb F-word across Genres and Time in DCOESL.

Based on Figure 1, it can be observed that the frequencies of lexical verb *f-word* in DCOESL are showing an ascending trend. In the 1960s and 1970s, *f-word* was not used at all in song lyrics across the four genres. However, in the 1980s, the *f-word* begins to emerge with only one occurrence. This occurrence increases to 8 in the 1990s, before it surges to 63 in the 2000s. The findings also revealed that the lexical verb *f-word* is most prevalent in the R&B genre with 30 counts pmw, followed by Rock and Pop with 24 counts pmw and 11 counts pmw respectively. Interestingly, there was no occurrence of the lexical verb *f-word* in the Country music genre. The common description of Rock music genre is that it is a genre of foul language and constant exposure to this genre has led to juvenile delinquency, undesirable attitude, stereotyped sex-roles and sexual violence (Marsh, 1993; Sutcliffe, 2011). The findings in this research have shown that this description is true to some extent. On the contrary, it was found that R&B contains higher *f-word*, than Rock and as well as Pop.

The study showed that the lexical verb *f-word* emerged in the 1980s, and continued to gain popularity in its use in the following two decades. At a glance of R&B history background, the genre begun from the American African community in the late 1940s. Ripani (2006) stated that R&B, a popular music genre primarily created by and for black Americans, prefers spoken song text and the lyrics contain slurred speech (as in f**k you), with implicit and explicit sexual contents especially since the 1980s (p.129). These explain the high preference of *f-word* for its adjacent collocates of personal pronouns I, you, me, and 'em (them), and the high counts of *f-word* itself in R&B. Tagged with USAS, R&B lyrics in DCOESL reflect mainly social context whereby intimate or sexual relationship is involved. This is because the genre is constantly associated with sexual topics (Hajdu, 2016). Based on the identified collocates of *f-word* in Table 6, f**k me and f**k 'em are highly significant to the

node f-word. The aforementioned collocations possess strong association, and also the collocates me and 'em are attested to be vital to *f-word* in song lyrics. The list of songs containing occurrence of these two collocations were retrieved from the diachronic corpus as shown in Table 4.

Table 4

List of Songs in DCOESL with F**k me and F**k 'em

No.	Collocation	No. of Occurrences (RF)	Artist	Song Title	Gender	Decade
	f**k me					
1		20	Ludacris	Splash Waterfalls	Male	2000s
	f**k 'em					
2		1	50 Cent	Ah!	Male	2000s
3		1	Ludacris	Move Bitch	Male	2000s
4		1	Jay-Z	I Just Wanna Love You	Male	2000s

Note. Collocations *f**k me* and *f**k 'em* are well attested for MI score and T-score. Thus, the two collocations have established a prominence feature as vital lexical verb *f-word* collocations in reflecting English language use in R&B song lyrics.

Collocation f**k me is mainly found in Ludacris' contemporary R&B (or Hip Hop) song; Splash Waterfalls. Meanwhile, collocation f**k 'em is found in three songs; 50 Cent's Ah!, Ludacris' Move B**ch, and Jay-Z's I Just Wanna Love You. The list illustrates two noticeable similarities in terms of gender and decade. All the song lyrics were sung by male black artists in the 2000s. The song lyrics contain both implicit (splash waterfalls) and explicit (the bitches want me to fuck 'em) sexual contents, with women as their objects. The hostility of black male singers towards black women in song lyrics can be closely linked back to the 1950s, the era of the Civil rights movement by the African Americans in the United States which lasted until year 1968 (Glasrud & Wintz, 2019). In general, this 15 years-long struggle was dedicated to attain equal rights and treatment of African Americans in the US, with many black female leaders, which brought up sexual politics into the lyrics of R&B (Ward, 1998, p. 71). The matriarchal nature of black society and the emasculation of black male contributed to the aggression towards the black women (Ward, 1998, p. 73).

This was explained by Moynihan (1995), whereby he traced the emasculation of black men to slavery when they were perceived as an incapable breadwinner and protector of their families (as cited in Ling & Monteith, 2014, p. 42). As Civil rights movement opened doors in job market for black women, black men struggled for the economic opportunity and political power to restore their patriarchy status. Black women were blamed for their political participations (Burrel, 2004, p.69). Otherwise

explained, the availability of job vacancies in the US for black women enabled them to be the functional head of their families. This caused resentments by the black males towards black women, which became male blues tradition to degrade black women for their 'quest for greater riches and better sex' in an attempt to exert 'male control' (Ward, 1998, p. 75).

Comparative Analysis of *F-word* in DCOESL with *F-word* in BNC and COCA

To further investigate on the use of *f-word* in DCOESL, comparative analysis with the reference corpora of BNC and COCA was carried out. This measure helps to identify which English language register does song lyrics most likely to mirror in general. Values for G^2 were calculated and tabulated in Table 5. The interpretation of results is based on the following two hypotheses by taking into account the significance values provided by McEnery et al. (2006, p. 55):

1. Null hypothesis, H_O: There exist no significant association between the occurrence of f**k in DCOESL with f**k in the reference corpora.
2. Alternative hypothesis, H_a: There exist a significant association between the occurrence of f**k in DCOESL with f**k in the reference corpora, $G^2>6.63$ at $p < 0.01$ or 1% level, $G^2> 10.83$ at $p < 0.001$, $G^2> 15.13$ at $p < 0.0001$.

Table 5

G^2 of F-word in DCOESL with F-word in Spoken and Written Registers of BNC and COCA

Subcorpora	RF	NF	G^2	Sig. Level
BNC$_{SPOKEN}$	106	10.18	54.61	<0.0001
BNC$_{WRITTEN}$	225	2.56	83.70	<0.0001
COCA$_{SPOKEN}$	4	0.01	102.40	<0.0001
COCA$_{WRITTEN}$	4266	7.55	62.74	<0.0001

G^2 test reveals that the difference between the corpora is highly significant. *F-word* in DCOESL is highly associated with *f-word* in the spoken register of reference corpora, $G^2_{BNCSPOKEN} = 54.61$ and $G^2_{COCASPOKEN} = 102.40$, df=1, p<0.0001. *F-word* in DCOESL is highly associated with *f-word* in the written register of reference corpora, $G^2_{BNCWRITTEN} = 83.70$ and $G^2_{COCAWRITTEN} = 62.72$, df=1, p<.0001. Based on the scores, the verb *f-word* in DCOESL shows highest resemblance to spoken register of the English language. To support the aforestated results, the researchers took a closer look at the emerging findings of adjacent collocations provided by the current diachronic corpus.

Adjacent Collocations of F-word

The top three adjacent collocations of the lexical verb *f-word* in DCOESL were calculated for the Mutual Information (MI) scores to assess the importance of the collocations to the node and show a clearer picture of the relationship between words (collocate + node) than that given by a simple collocation list alone. For BNC and COCA, the MI scores for *f-word* collocations are retrieved from their freely accessible websites.

To be certain that the collocations are the results of more than vagaries, another collocation measurement was calculated; the T-score. To avoid misleading results from the MI scores, T-score is utilised to analyse and validate MI scores by giving clearer insight to which words have a strong attraction to the lexical verb *f-word* and which do not occur frequently in DCOESL are not given high significance. Table 6 shows the MI and T-score values for *f-word* collocations in DCOESL. The interpretation of results for MI is based on the following two hypotheses by taking into account the significance values provided by Hunston (2002, p. 71):

1. Null hypothesis, H_O: *F-word* and the adjacent collocate occur randomly with each other, MI < 3.

The interpretation of results for T-score is based on the following two hypotheses by taking into account the significance values provided by Hunston (2002, p.72):

1. Null hypothesis, H_O: The adjacent collocate is not significant to the node *f-word*, T-score < 2.

Table 6

MI and T-score Values for F-word Collocations in DCOESL (Window Span: -1, +1)

Rank	-1	RF	NF	MI	T	+1	RF	NF	MI	T
1.	I	11	8	0.369	0.638	me	18	13	3.442	3.274
2.	you	9	6	0.042	0.071	it	7	5	2.124	1.723
3.	can't	3	2	2.261	1.119	'em	6	4	7.078	1.985

As shown in Table 6, the collocate first person singular subjective personal pronoun I occurs most frequent directly to the left of *f-word* with 8 occurrences pmw. However, collocate I has an MI score of 0.369 and a T-score as low as 0.638. This means that I, is not a significant collocate and it has very weak attraction to *f-word*. The second most frequent collocate is the second person personal pronoun you with MI score of 0.042 and T-score of 0.071. Although identified at rank 2, you is not a significant and strong collocate of *f-word*. Similar goes to can't and it in terms of collocational strength and collocates significance. Table 7 shows the top three most common word forms in DCOESL.

Table 7

Top Three Most Common Word Forms in DCOESL

Rank	Word	RF	NF	%
1	I	66637	47173	4.717
2	you	57569	40754	4.075
3	the	45214	32008	3.201

Personal pronouns I and you occupy 4.717 per cent and 4.075 per cent respectively of the entire DCOESL. In other words, the words I and you are very frequent in the English language as a whole. They tend to occur near the top of many collocate lists simply because it is so frequent overall. Likewise, personal pronouns I and you are very common in BNC and COCA. Table 8 shows the distribution of top two pronouns in overall of BNC and COCA and Table 9 shows the distribution top two pronouns in subcorpora of BNC and COCA.

Table 8

Distribution of I and You in Overall of BNC and COCA

		Corpora				
		BNC		COCA		
Word	Rank	NF	%	Rank	NF	%
I	14	8660	0.9	11	9797	1
you	19	6787	0.7	14	7869	0.8

Table 9

Distribution of I and You in Subcorpora of BNC and COCA

	Subcorpora							
	BNC$_{Written}$		BNC$_{Spoken}$		COCA$_{Written}$		COCA$_{Spoken}$	
Word	NF	%	NF	%	NF	%	NF	%
I	6223	0.6	29241	2.9	6650	0.7	3147	0.3
you	4538	0.5	25780	2.6	4148	0.4	3720	0.4

Based on Table 8 and Table 9, the personal pronouns I and you are both in the top 20 of common word forms of BNC and COCA. COCA has the higher occurrences of I and you, with 9797 and 7869 occurrences respectively. For BNC, I and you in the spoken register are higher than written register by 2.3 percent and 2.1 percent respectively. On the other hand, COCA, I in the spoken register is lower than written register by 0.4 percent. You has about the same occurrences, which amount to about 0.4 percent of total occurrences in both registers.

The third most frequent collocate to the left of *f-word* is contraction can't, which has an MI score of 2.261 and T–score of 1.119. The scores are not adequate

enough (MI < 3 and T < 2), to prove that can't is a highly significant collocate of fuck. In DCOESL, can't amounts to 2820 occurrences overall in DCOESL. Figure 2 shows a concordance set of can't fuck in DCOESL**.**

1	can't fuck with this And niggas *can't fuck*	with that Shit that I drop 'cause ya know
2	Y'all niggas *can't fuck*	with the (Boy) I 'm tellin ya (Boy) Put
3	nigga Mr Dr. Dre Like I said , niggas *can't fuck*	with this And niggas can't fuck with

Figure 2. Concordance Set for Can't F**k (Pronoun+Verb$_{Base}$) in DCOESL.

For window span +1, the collocate objective personal pronoun, me, shows significant MI score of 3.442 and T-score of 3.274. The aforementioned scores indicate that me is an important collocate to *f-word*. The T-score reveals that the occurrences of f**k me are not purely due to chance. In other words, we can be certain that the association of f**k me is highly non-random. The reduced form of third person plural objective personal pronoun them, which is 'em, has a high MI score of 7.078. The T-score is 1.985, which is not high enough to prove that 'em is a strong collocate. This means that the association between collocate 'em and node *f-word* are highly non-random. However, the T-score highlighted that 'em is not strongly associated to the node *f-word*. Figure 3 shows a set of concordance lines for colocation f**k me. F**k me occurs in R&B with 16 counts (lines 2 until 18) and Rock with one count (line 1) within the 2000s. F**k me does not occur in Country and Pop.

1	You 're crazy, But I like the way you *fuck me*	
2	Just scratches on your back, what? (*fuck me*	!) Ex's ain't acting right, and you
3	Then hear her scream your name, what? (*fuck me*	!) Follow this DICK-tionary, you're
4	Range Rover all wood Do me like you should, *fuck me*	good, suck me good
5	Say it (make love to me) Oh! Oh! What? (*fuck me*	!) I 'm bout to throw some game
6	and every night You got to do it right, what? (*fuck me*	!) I 'm bout to throw some game
7	You call her Jabberjaws, what? (*fuck me*	!) I 'm bout to throw some game
8	Y'all just don't give a fuck, what? (*fuck me*	!) I 'm bout to throw some game
9	she like it rammed up, Ropes and handcuffs, what? (*fuck me*	!) I 'm bout to throw some game
10	on the low, And meet 'em right at for, what? (*fuck me*	!) Nothing but fights and fussing
11	and smacking ass, You got to make it last, what? (*fuck me*	!) Together holding hands, you out
12	down to ya fingertips Trying that freaky shit, what? (*fuck me*	!) Turn on some Babyface, just for
13	now how to drive a stick, what? (*fuck me*	!) You both unite as one, you the
14	every night, you got to do it right, what? *fuck me*	You 'bout to buy a ring, she needs the
15	all the while Doggy and froggy style, what? (*fuck me*	!) You in between the sheets, licking
16	and every night You got to do it right, what? (*fuck me*	!) You wanna tell the world, cause she 's
17	and every night You got to do it right, what? (*fuck me*	!)
18	and every night You got to do it right, what? *fuck me*	They want it nice and slow, kiss 'em

Figure 3. Concordance Set for F**k Me (Verb$_{Base}$ + Pronoun Word Order) in DCOESL.

The collocate impersonal pronoun, it, has an MI score of 2.124 and T-score of 1.723. The aforestated scores are not high enough (MI < 3 and T < 2) to prove that it is a significant collocate. In other words, it has not a very strong association and attraction to the node fuck. Figure 4 below shows a set of concordance lines of f**k it.

1	she can get it too. *Fuck it*	though, I'm honest yo I'm saying
2	get this money and act like a damn fool. *Fuck it*	, act stupid and jump in the damn pool
3	She like *fuck it*	, I 'll just stay down here and do hair
4	Might light your wrist, but that about it. *Fuck it*	, I might wife you and buy you nice whips
5	I ain't even goin' act holier than thou, Cause *fuck it*	, I went to Jacob with 25 thou
6	or I 'm caught in "The Matrix" But *fuck it*	, let the Fish-burne Red or green pill
7	Well *fuck it*	, Van Damme 'em

Figure 4. Concordance Set for F**k It (Verb$_{Base}$ + Pronoun Word Order) in DCOESL.

The reduced form of third person plural objective personal pronoun them, which is 'em, has a high MI score of 7.078. The T - score is 1.985, which is not high enough to prove that 'em is a strong collocate. This means that the association between collocate 'em and node fuck are highly non-random. However, the T-score highlighted that 'em is not strongly associated to the node fuck. Figure 4 below shows a set of concordance lines of fuck 'em.

1	With no regard for the law, we dodge 'em like, "*fuck 'em*	all" But I won't get caught up
2	she screamed unto me, A free for all *Fuck 'em*	all You're on your side
3	through black and white War horse, warhead *Fuck 'em*	all man, white knuckles tight
4	nigga fresh out the, slum Automatic, gun *Fuck 'em*	one-on-one We wrap up ya punk ass
5	successful a sex symbol The bitches want me to *fuck 'em*	true Hold up, wait up, shorty
6	in the system, ain't no tellin' Will I *fuck 'em*	will I diss 'em, that's what they be

Figure 5. Concordance Set for F**k 'em (Verb$_{Base}$ + Pronoun Word Order) in DCOESL.
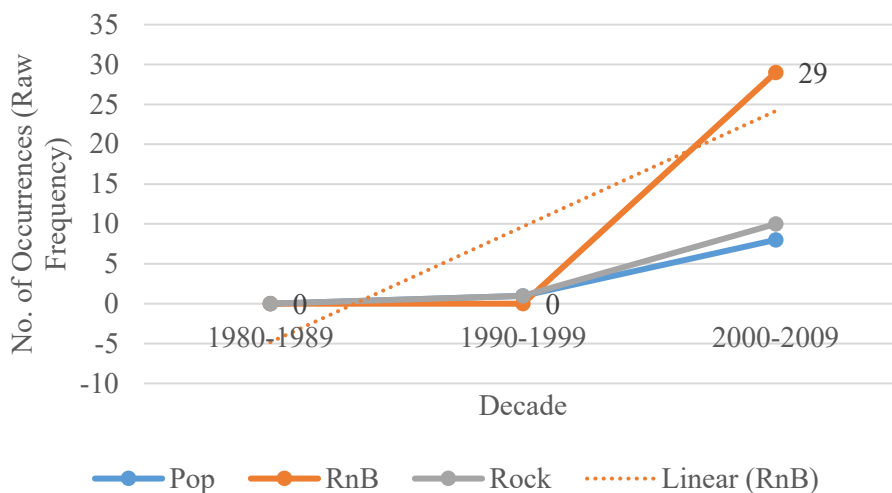

Figure 6. Distribution of *F-word* + Pronoun across Genres and Time in DCOESL.

Based on Figure 6, *f-word* begins to emerge in the 1980s, particularly in R&B. Note that the attested occurrences of lexical verb *f-word* in DCOESL are mostly combined with personal pronouns such as me and 'em. This attribute of extremely

common occurrences of personal pronouns and emotive verbs (*f-word* is categorized as verb of emotion via USAS) resembles spoken English, which Biber et al. (1999) described as the characteristics of conversation register. The spoken-like characteristics of R&B through the use of lexical verb *f-word* marks the beginning of transition in vocal style from gospel in early 1980s, to spoken word which became very prominent by late 1990s (Burnim & Maultsby, 2014, p. 266). This explains the surges of *f-word* + Personal Pronouns phenomena in DCOESL as can be seen in Figure 6.

Out of 104 sentences in the lyrics containing the lexical verb *f-word*, this pattern covers a total of 41 occurrences or 39.42 percent of the sentences. This pattern is found to be significant when the top three collocation of *f-word* was generated via AntConc. The use of personal pronouns including I and you has been traditionally associated with informality (Wales, 1996, p. 107). The Personal Pronoun + VerbLexical construction is associated with informal talk (Trillo, 2008, p. 71), to express informality and speaker's personal involvement, which is typical of informal registers such as fiction and informal speech (Březina, 2018, p. 164) as can be found in BNC and COCA. For DCOESL, I and you are frequently used for their significance in illustrating strong personal involvement of the artists with the addressees as they convey their stories through song lyrics, and make the audience feel like they "participate" in those stories (Griffee, 1992, p. 4). Semantically tagged, I f**k and you f**k in DCOESL generally refer to artists' intimate or sexual relationship, human anatomy and physiology, and carefree attitude.

The phrase 'f**k with' as in 'f**k with somebody' means 'mess with' which means to 'treat somebody badly in a way that makes them annoyed' (Hornby et al., 2010, p. 605). In DCOESL, can't *f-word* means 'can't mess' with something or somebody. The contraction or negation can't is a sign of colloquialization and it is very common in non-academic language (Iosef, 2013; Biber et al., 1999; Biber et al., 2002). This is true in the case of COCA whereby can't is the highest with 593pmw in written fiction register, but lowest with 55 occurrences in academic writing register. For BNC, can't amounts to 1239 occurrences in spoken register and 197 occurrences in written register. Thus, it can be said that DCOESL contains high use of negative contraction can't and its combination with taboo word *f-word* (Negation + VerbBase: can't f**k) proves that DCOESL is a written genre which contains informal speech-like feature.

If f**k me was to be interpreted according to literal meaning, it could be defined as a request from someone to have sexual intercourse with him or her. However, this is not the case for *f-word* me in DCOESL. Table 10 shows semantic categories of *f-word* in DCOESL.

Table 10

Semantic Categories of F-word in DCOESL

Word	POS Tag	USAS Semantic Tag	Category		
f-word	VV0	S3.2/B1 E6+	S	:	Social actions, states and processes.
			S3.2	:	Relationship: Intimate/sexual
			B1	:	Anatomy and physiology
			E	:	Emotion
			E6+	:	Confident

Semantically tagged using the USAS English tagger, *f-word* mainly refers to social action, states and processes, particularly intimate or sexual relationship. It also refers to anatomy and physiology. Interestingly, *f-word* as in f**k me also refers to positive emotion, that is to express confidence. This could be black men's way of exercising their masculinity power. This finding is contrast to the definition given by Stuart-Hamilton (2007), that is to express surprise, although he also noted that 'f**k me' is not to be taken literally (p. 94).

Conclusion

Computational Linguistics provides great possibilities for linguistic investigations as compared to traditional research methods. The corpus computational tools have become indispensable as automatic linguistic annotation, flexible query and quantitative and qualitative analysis of large text corpora can be conducted to gain deeper understanding of the natural language. The lexical verb *f-word* in DCOESL is proven as a marker for informal spoken English. In term of statistical test of significant against reference corpora, *f-word* resembles the spoken register of COCA the most. Its frequent collocates are consisting of pronouns I, you, me, it, and 'em, and colloquial language signs; contractions such as can't and 'em. From the semantic tagging applied to DCOESL, *f-word* in song lyrics generally reflect social actions, states and processes. This taboo verb is very prominent in the RnB genre because lyrics of Rhythm and Blues never stray far from the topic of sex (Hajdu, 2016). The history background of *f-word* in R&B song lyrics could be traced back to the Civil rights movement which caused black men's hostility towards black women. From the findings in DCOESL, lexical verb *f-word* does not necessarily to be defined by its literal definition, but also can be interpreted as an expression of confidence namely f**k it, f**k 'em, and f**k me.

Funding Acknowledgment

References

Bartle, G. (1962). Music in the language classroom. *Canadian Modern Language Review*, *19*(1), 11-14. https://doi.org/10.3138/cmlr.19.1.11

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. Pearson Education Limited.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finnegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.

BNC Consortium. (2018). British National Corpus: BNCweb (CQP-Edition). http://bncweb.lancs.ac.uk/cgibinbncXML/BNCquery.pl?theQuery=search&urlTest=yes

Březina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Bridle, M. (2018). Male blues lyrics 1920 to 1965: A corpus based analysis. *Language and Literature, 27*(1), 21-37. https://doi.org/10.1177/0963947017751757

Burnim, M. V. & Maultsby, P. K. (2014). *African American music*. Routledge.

Burrel, B. C. (2004). *Women and political participation: A reference handbook.* ABC-CLIO.

Domoney, L. & Harris, S. (1993). Justified and ancient: Pop music in EFL classrooms. *ELT Journal, 47*, 234-241.

Eken, D. K. (1996). Ideas for using pop songs in the English language classroom. *English Teaching Forum*, *34*, 46-47.

Glasrud, B. A. & Wintz, C. D. (2019). *Black Americans and the civil rights movement in the west.* University of Oklahoma Press.

Goddard, C. (2014). Jesus! vs Christ! In Australian English: Semantics, secondary interjections and corpus analysis. In J. Romero-Trillo (Eds.), *Yearbook of corpus linguistics and pragmatics 2014: New empirical and theoretical paradigms* (pp. 55-77). Springer International Publishing.

Griffee, D. T. (1992). *Songs in action.* Phoenix ELT.

Hajdu, D. (2016). *Love for sale: Pop music in America.* Farrar, Straus and Giroux.

Hornby, A. S., Turnbull, J., Lea, D., Parkinson, D., Phillips, P., Francis, B., Webb, S., Bull, V, & Ashby, M. (2010). *Oxford advanced learner's dictionary international student's edition (8th ed.).* Oxford University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Iosef, M. (2013). *Signs of colloquialization: Three corpus-based case studies* [Master's thesis, University of Oslo). University of Oslo Library. https://www.duo.uio.no/bitstream/handle/10852/37034/iosef_master.pdf?isAllowed=y&sequence=2

Jolly, Y. (1975). The use of songs in teaching foreign languages. *The Modern Language Journal, 59*(1,2), 11-14. https://doi.org/10.2307/325440

Krashen, S. T. (1982). *Principles and practice in second language acquisition*. Pergamon.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

Ljung, M. (2011). *Swearing: A cross-cultural linguistics study.* Palgrave Macmillan.

Lo, R. & Fai Li, H. C. (1998). Songs enhance learner involvement. *English Teaching Forum*, *36*(3), 8-11.

Logan, B., Kositsky, A., & Moreno, P. (2004). *Semantic analysis of song lyrics*. Paper presented at IEEE International Conference on Multimedia and Expo, Taipei, Taiwan.

Marsh, D. (1993). *Louie Louie: The history ofmythology of the world's most famous Rock 'n' Roll song.* Hyperion.

McEnery, A. M. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present.* Routledge.

McEnery, A. M., & Xiao, Z. (2003). F-word revisited. In D. Archer, P. Rayson, A. Wilson, & A. M. McEnery (Eds.), *Proceedings of the corpus linguistics 2003 conference* (pp. 504-512). Centre for Computer Corpus Research on Language Technical Papers, Lancaster University.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book.* Routledge.

Motschenbacher, H. (2016). *Language, normativity, and Europeanisation: Discursive evidence from the Eurovision Song Contest.* Palgrave Macmillan.

Moynihan, D. P. (1965). U.S. Department of labor. The Negro family: The case for national action. In P. J. Ling & S. Monteith (Eds.), *Gender in the civil rights movement* (pp. 42). Garland Publishing, Inc.

North, A. C., Hargreaves, D. J., Hargreaves, J. J. (2004). The uses of music in everyday life. *Music Perception*, *22*, 63–99.

Petrie, K. J., Pennebaker, J. W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts, 2*(4), 197-202.

Richards, J. (1969). Songs in language learning. *TESOL Quarterly*, *3*(2), 161-174. https://doi.org/10.2307/3586103

Ripani, R. J. (2006). *The new Blue music: Changes in Rhythm and Blues (1950-1999).* University Press of Mississippi.

Saarinen, E. (2013). *From war pigs to unsung heroes: The criticism and justification of war in metal lyrics* [Master's thesis, University of Turku). Turun Yliopisto. http://urn.fi/URN:NBN:fi-fe201306123973

Schoepp, K. (2001). Reasons for using songs in the ESL/EFL classroom. *Internet TESL Journal*, *2*(2). http://iteslj.org/Articles/Schoepp-Songs.html

Sinclair, J. M. (2015). *How to use corpora in language teaching.* John Benjamins Publishing Company.

Stuart-Hamilton, I. (2007). *An Asperger dictionary of everyday expressions.* Jessica Kingsley.

Sutcliffe, P. (2011). *AC/DC: High voltage Rock 'n' Roll the ultimate illustrated history.* Voyageur Press.

Taina, J. (2014). *Keywords in heavy metal lyrics* [Master's thesis, University of Helsinki). Helda. https://helda.helsinki.fi/bitstream/handle/10138/136524/keywords.pdf

Tognini-Bonelli. (2001). *Corpus linguistics at work.* Amsterdam.

Trillo, J. R. (2008). *Pragmatics and corpus linguistics: A mutualistic entente.* Mouton de Gruyter.

Wales, K. (1996). *Personal pronouns in present-day English.* Cambridge University Press.

Ward, B. (1998). *Just my soul responding: Rhythm and Blues, black consciousness and race relations.* UCL Press.

Biography

Flora Goyak holds a Master's degree in TESL from Sultan Idris University of Education. Her specializations are Corpus Linguistics, Computational Linguistics and TESL.

Mazura Mastura Muhammad is a senior lecturer at the Department of English Language and Literature, Universiti Pendidikan Sultan Idris and her areas of specialization are Corpus Linguistics, Language Assessment and TESL.

Muhammad Fadzllah Zaini is a lecturer in the Department of Malay Language and Literature, Universiti Pendidikan Sultan Idris (UPSI). His areas of specialization are Corpus Linguistics and Computer Applications in Language Education.

Wesam Mohamed Abdelkhalek Ibrahim is a senior lecturer at the Department of Basic Sciences, Community College, Princess Nourah bint Abdulrahman and the Department of Foreign Languages, Faculty of Education, Tanta University, Egypt.

Ardi Gunsuh holds a Bachelor's Degree in International Business majoring in International Finance from Universiti Malaysia Sabah and is an integral figure within TalentCorp's Graduate and Emerging Talent department where he focuses on campus engagement activities, actively engaging with the public and private institutes of higher learning. He co-leads TalentCorp's Structured Internship Programme, collaborating with employers to ensure that Malaysia's graduates are ready for the workforce. Since joining TalentCorp in 2013, Ardi has contributed to various departments, including the Industry Partnership team where he established meaningful partnerships with key industry players to benefit Malaysia workplace and workforce via TalentCorp's initiatives.