



Linnæus University

Sweden

Bachelor thesis

We will rock you

*A diachronic corpus-based analysis
of linguistic features in rock lyrics*



Author: Johanna Falk
Supervisor: Magnus Levin
Examiner: Diane Pecorari
Semester: Autumn 2012
Course code: 2EN10E

Abstract

In the present paper, the potential of corpus-linguistic research is put into action. More specifically, a corpus-based demonstration of the general style used in rock lyrics is formed in order to identify the genre-specific features. A corpus consisting of roughly 53 000 words was created for this research. The focus lies on a quantitative and qualitative analysis of the vocabulary as well as of the stylistic markers. The aim of this essay is to investigate the language used in the lyrics of rock music lyrics and the results of the research illustrate in what ways rock lyrics are either more spoken-like or written-like; whether rock lyrics mirror the general word usage in society, and in what ways rock music is comparable to other genres. Using the rock lyrics corpus (ROLC), trends within rock lyrics were retraced diachronically. Results show that rock lyrics, to some extent, follow the general word usage. However, other results also contradict this. Further research in this area is therefore encouraged. Findings show that rock music shares features with other genres, but also that that rock has some unique features. Developments and stagnations were detected in regards to word usage. These features are examined in order to find an explanation.

Keywords

Corpus research, lyrics, rock music, genres, diachronic, AntConc, word usage

Table of contents

1.	Introduction.....	1
1.1	Aim, scope and research questions	2
2.	Theoretical background	3
2.1	The definition and creation of a corpus	5
2.2	Previous corpus-based studies	7
3.	Material and Method.....	8
3.0	Defining rock	8
3.1	Compiling a corpus	9
3.2	Difficulties in compiling a corpus.....	10
4.	Results.....	12
4.1	Comparing genres	14
4.2	Interjections	17
4.3	Non-standard spellings	18
4.4	Comparing fields.....	19
4.5	Diachronic results	21
5.	Conclusion	25
	References	

1. Introduction

“I’m a fool but I’ll love you dear until the day I die

Now and then, there’s a fool such as I”

– Elvis Presley, “A fool such as I” (1959)

“And I would do anything for love,

I’d run right into hell and back”

– Meatloaf, “I’d do anything for love” (1993)

Rock music abruptly found its way into music history, and consequently into human history, in the early 1950s, owing to artists such as Elvis Presley, Chuck Berry and Bill Haley who thus helped initiate the rock movement. However, the answer to whether it is the tune or the lyrics which have helped this confident genre to take a hold on modern music might never be justly revealed. Pettijohn & Sacco argue that:

“Lyrics are an important form of communication, serving a variety of purposes as documented in the psychology of language literature. Lyrics tell stories and communicate with audiences in a manner similar to how people have conversations with each other [...] When accompanied with words, songs allow us to communicate emotions, tell stories, and even express our opinions and attitudes” (2009:297).

However, are there any words and themes within the rock genre that have a history of always being particularly fashionable? If that is the case, the following question is therefore whether there are any reasons for their popularity and, for why certain words gained their popularity at a certain time. By analyzing lyrics, interesting trends in language may be uncovered.

However, the reasons and circumstances behind them must be investigated further to reach a linguistically valuable conclusion. Pettijohn & Sacco claim that “when social and economic times are more threatening, Billboard songs with more words per sentence, more future references, and more coverage of social processes and other references are popular” (2009: 307). Individuals' music preferences are both a subconscious reflection of specific personality characteristics, but also a representation of their beliefs, attitudes and needs, , according to Peter J. Rentfrow & Samuel D. Gosling (2006), Consequently, given that music reflects

individuals and individuals evidently are members of society, should not music, in turn, reflect society and vice versa?

1.1 Aim, scope and research questions

The aim of this essay is to investigate how language is used in rock music lyrics and whether the word usage in rock music acts as a contemporary mirror of the overall word usage in society. In addition, particular features of rock music and its lyrics will also be addressed. Furthermore, word usage in the rock genre will be compared with the word usage in other genres. This essay will include a summarizing rock lyrics corpus (henceforward referred to as ROLC) corpus of rock lyrics, ranging from the 50s to the 90s, where the lyrics of 60 songs from each decade are included. These lyrics will be subjected to both quantitative and qualitative analysis. The aim will be addressed by answering the following research questions:

- What are the linguistic features of rock lyrics, which words are the most frequent and what are the possible reasons for these findings?
- Are lyrics of the rock genre comparable with other genres?
- Are there any linguistic similarities between the word usage in rock lyrics and the overall word usage in society, or to put it simply – does the language of rock lyrics mirror that of society?
- Has the usage of certain words changed over time or is it still the same?

2. Theoretical background

Examining the lyrics of rock music is a research rather unique to this essay. However, in a certain sense, the topic chosen for this essay has been approached before. In their article, Pettijohn & Sacco (2009) investigated the lyrical content of Billboard No. 1 songs, reaching from 1955 to 2003, by means of the software LIWC (Linguistic Inquiry and Word Count). They came to the conclusion that there was a trend towards more meaningful, comforting and romantic lyrics during more threatening socio-economic times¹. In Kreyer & Mukherjee's (2007) article the analytic potential of corpus-linguistic research in the genre of pop song lyrics was discussed. Furthermore, Cole (1971) analyzed the lyrics of the annual top 10 single songs during each year of the 1960s. However, Cole analyzed the overall emotion of the lyrics in these particular songs while this essay will explore the matter of linguistic features, such as interjections and non-standard spelling. Furthermore, this essay will not research in the genre of top *pop* song lyrics but, as mentioned, instead investigate the genre of *rock* song lyrics. This kind of extensive research on the subject of rock song lyrics alone has not been executed earlier. Nevertheless, Katznelson, Gelman, Lindblom & Caput come relatively close:

“The sociolinguistic approach to this study included analyzing frequency lists in each genre to find highly frequent words both in each genre, and across genres [...] The main sociolinguistic research questions asked were: What words are highly frequent in each genre? Is there a lot of overlap between genres? Do these words fit with the stereotypes about each genre? The main pedagogical questions asked were: What can music lyrics show us about the phraseologies of a specific lexical item and how can this type of corpus-based investigation be used to facilitate both the teaching and learning of English?” (2010:2)

As in the current essay, Katznelson et al.'s corpus-based research concerns word frequency. However, in their research they focus on four genres of American popular music – pop, rock, country and hip hop – rather than one separate genre. Furthermore, this essay does not pay any attention to the pedagogical aspects of lyrics since it focuses on the linguistic aspect only. The corpus-based study in Katznelson et al.'s article extends over a period of the past twenty

¹ A questionnaire was used to assess the overall emotion of the lyrics. In the questionnaire a 7-point Likert-type scale ranging from 1 = strongly disagree to 7 = strongly agree was answered by fifty-four lyric raters.

years whereas this essay range from the 50s to the 90s (however, as mentioned, one then has to keep the four different genres in mind). The study in Katznelson et al.'s article is based on 110 rock songs, 102 pop songs, 112 country songs and 109 hip hop songs. Thus a total of 433 songs and 178,982 words were collected. The time span and overall size “were chosen to ensure representativeness of each genre; thus minimizing any idiosyncratic features.” (2010:6). Each researcher was given one genre to collect data from, and each genre was selected “by referring to the top musical categories from the Billboard charts” (2010:6). The results are shown in table 1.0 below.

Table 1.0 The ten most frequent nouns by genre (based on Katznelson et al.'s article (2010:12))

Rock	Pop	Country	Hip hop
Love	Love	Love	Life
Time	Baby	Way	Time
Way	Way	Heart	Ass
Pain	Girl	Baby	Baby
World	Boom	Night	Thing
Life	Heart	Time	Way
Baby	Time	Man	Girl
Eyes	Life	Life	Music
Head	World	Day	Bitch
Heart	Man	Thing	Money

In table 1.0, cultural ideations, as Katznelson et al. (2010) phrase it, reveal themselves. The frequency/distribution within each genre “shows the type of influence they enact upon their audience [...] the nouns in rock generally reflect the heartbreak type of love we hypothesized versus the carefree love of pop.” (2010:14)

However, perhaps the more interesting results are the ones showing words unique for each particular genre:

Table 1.1 Genre unique words

Rock	Pop	Country	Hip Hop
Pain .61	Boom .95	Night .77	Ass 1.04
Eyes .44		Day .62	Music .67
Head .41			Bitch .66
			Money .62

Katznelson et al.'s study is reminiscent of the sort of research that will be carried out in this essay in several ways, with the exception that the focus in the current essay will be on the rock genre. Moreover, the results are based on diachronic research, not synchronic.

DeWall, Pond, Campbell & Twenge's study (2011) attempt to determine whether lyrics in popular U.S. songs changed over time. However, their focus was on manner that mirrored documented psychological changes across the same time period, and even though it was mostly psychologically oriented, their study contributed to the current essay. Their research claims that "cultural products can also be used to understand psychological changes *within* a culture over time, exploring whether the linguistic contents of popular song lyrics change over time in tandem with generational shifts in personality traits over the same period of time" (2011:201). According to DeWall et al, feelings of loneliness and social isolation rose 250 per cent between 1985 and 2004.

2.1 The definition and creation of a corpus

When creating a corpus, the road to perfection is at times uncertain. However, an insight into previous authors' works can contribute to a desirable outcome. One of these previous works is Petrie, Pennebaker & Sivertsen's study and the methods used in their research have been partly embraced in this essay, as will be shown in the material and method section. Firstly, they downloaded the lyrics of Beatles' songs containing more than 50 words. Their final database included 185 songs which were converted into conventional American English spelling. Phrases or choruses that were repeated three times or more within a song were deleted. Having done that, each song was converted into an individual text file. The result of their research showed that:

“songs from the early years of the Beatles were characterized by positive emotion with many tunes concerning the joys of new romantic relationships. Their early

records maintained a largely positive emotional tone [...] The emotional content of songs continued after this period to be more psychologically distant and melancholic and less positive with many songs looking back to happier times” (2008:200).

To be able to create a corpus, one must first know what a corpus is. As of yet, there is no proper manual for corpus creation. However, in their article, Atkins, Clear & Ostler (1991) provide basic guidelines and definitions for composing a corpus. They start off by defining text collections, and a units of text, into four different types of text collection: archive, electronic text library (or ETL), corpus, and sub corpus. Their focus, and also this essay's focus, lies on the corpus type. Since corpora consist of texts, the issue of finding a definition of what a text is is soon raised. The characteristics given are: “it is discursive and typically at least several pages long; it is integral; it is the conscious product of a unified authorial effort; it is stylistically homogeneous” (1991:2). However, since this essay deals with a specific kind of text these characteristics are not accurately followed since a song lyric is, in most cases, not several pages long nor conversational; it stands on its own and is not part of a longer text; and it may have multiple authors. Moving on, the article by Atkins et al. also brings up the planning stage: both the linguistic design of the corpus but also costs and administration (this will not be addressed here). Certain basic tools like word frequency and concordancing will have to be handled, as will more advanced text handling like collocation and lemmatization. While *Corpus Design Criteria* acts as a good template, it caters to more advanced studies.

Nesselhauf (2011) also tries to answer the question of what a corpus is in her article She argues that “a corpus can be defined as a systematic collection of naturally occurring texts” (2011:2). That is, the arrangement and contents of the corpus follow certain extralinguistic principles. She continues:

“For example, a corpus is often restricted to certain text types, to one or several varieties of English, and to a certain time span. If several subcategories (e.g. several text types, varieties etc.) are represented in a corpus, these are often represented by the same amount of text. “Systematic” also means that information on the exact composition of the corpus is available to the researcher (including the number of words in each category and in the whole corpus, how the texts included in the corpus were sampled etc.) Although “corpus” can refer to any systematic text collection, it is commonly used in a narrower sense today, and is often only used to refer to systematic text collections that have been computerized” (2011:2).

The article lists a number of different corpora and also a couple of corpus softwares. With these types of corpus softwares one might search for the occurrence of certain strings and concordance–lines, sort and also narrow down the results, etc.

2.2 Previous corpus–based studies

In their article Schneider & Miethaner (2006) introduce their corpus BLUR (short for Blues Lyrics collected at the University of Regensburg), which is a corpus of Blues lyrics, compiled in 2006 at the University of Regensburg in Germany. It comprises over 8000 Blues lyrics from the early twentieth century and its purpose is to compile and examine the roots of African American Vernacular English, through blues. The design of BLUR is presented and practical consequences resulting from the nature of the texts are discussed. Schneider & Miethaner argue that blues lyrics can be a valuable source of linguistic data.

Several previous studies have been intuition–based, not corpus–based like Schneider & Miethaner's research. In their research they could therefore point out and focus on previously undetected syntactic constructions in a variety of English (AAVE). With BLUR as a linguistic corpus they have found a discovery procedure and one of their discoveries, the *start to V-ing* pattern, are analyzed in detail in BLUR and compared with other electronic corpora. Through this they can present evidence of possible origins and at the same time achieve their goal of illustrating the value of BLUR in particular and of historical nonstandard language in general.

In Kreyer & Mukherjee's article (2007) they examine the stylistic variation in pop song lyrics through corpora: "To this end, we envisage the compilation, annotation and analysis of a large and representative corpus of pop song lyrics, the Giessen–Bonn Corpus of Popular Music (GBoP). In the present paper, we report on general findings from a pilot investigation of the pilot version of GBoP. The focus here is on the identification of style markers of pop song lyrics in general and on a quantitative and qualitative analysis of the vocabulary and of lexicogrammatical routines in the corpus." (2007:31). In its pilot form, the Giessen–Bonn Corpus of Popular Music includes 27 of the top 30 albums in the US Album Charts from 2003, for a total of 442 songs and approximately 176,000 words. As mentioned, in their paper they use both quantitative as well as qualitative analysis, but in contrast to the present essay, the GBoP only covers songs from a single year.

3. Material and Method

3.0 Defining rock

To be able to collect the right material the source of the search has to be defined first. What is rock music? To answer that question one first has to know where it originated from. Rock, or rock'n'roll, is a form of popular music, which emerged in the United States in the 1950s, and soon spread all over the world. It is truly difficult to define rock since it has been divided into such a broad variety of subgenres like hard rock, soft rock, folk rock, industrial rock, metal and the list goes on. Important to point out is that in this essay *rock* acts as a generic term for all subgenres evolved out of *rock 'n' roll*.

Even Encyclopedia Britannica has problems defining rock:

There is basic agreement that rock “is a form of music with a strong beat,” but it is difficult to be much more explicit. The Collins Cobuild English Dictionary, based on a vast database of British usage, suggests that “rock is a kind of music with simple tunes and a very strong beat that is played and sung, usually loudly, by a small group of people with electric guitars and drums,” but there are so many exceptions to this description that it is practically useless. Legislators seeking to define rock for regulatory purposes have not done much better. The Canadian government defined “rock and rock-oriented music” as “characterized by a strong beat, the use of blues forms and the presence of rock instruments such as electric guitar, electric bass, electric organ or electric piano.” This assumes that rock can be marked off from other sorts of music formally, according to its sounds.

In de Clercq & Temperley's analysis (2011) of rock harmony they address the issue of definition as follows:

“Clearly, there is some disagreement as to what is rock and what is not. One might surmise that rock, or rather the ‘rock song’, is not a discrete, well defined category, but a loosely defined schema, of which some songs are very clear-cut instances and others are more borderline and ambiguous. It is also possible, of course, that different people simply have different understandings of the term. In light of this, and given the aim of the study to employ empirical, statistical methods, it seems most sensible to adopt a statistical approach to the definition of rock as well, that is, basing it on the opinions of many individuals” (2011:51).

Thus, according to Clercq & Temperley, rock music is defined by the opinion of the majority.

For lack of a proper definition, perhaps we have to resort to thinking in terms of prototypes². Examples of such prototypes are artists/bands like Elvis Presley (3% of the corpus material in this research) – who figured a great deal during the 1950s –, The Beatles and The Rolling Stones (about 7% of the corpus material) – popular in the 1960s and 1970s –, Queen, U2 and Prince (about 3% of the corpus) – popular during the 1980s – and Nirvana and Red Hot Chili Peppers (about 2% of the corpus) – popular in the 1990s.

Not known for being the most reliable source of information, Wikipedia has nevertheless proven to be a helpful “key”. The website provides every song with a classification, which has helped rule out the borderline cases. A good example of an artist balancing on the borderline between pop and rock was the artist Michael Jackson. Even though he was known as the King of Pop he at times was on the verge of rock. His song *Beat It* has been included in this corpus. Another borderline case is the group *Green Day* which at times border between rock and punk. Their song *Minority* was not included in the corpus since it was defined as pop punk according to Wikipedia.

3.1 Compiling a corpus

The quantitative part, and the foundation of this essay, consists of a corpus including rock lyrics from the 1950s and onwards to and including the year 1999. The reason for this time span is to make an investigation of diachronic trends possible. The corpus contains the most popular titles from each year and decade. Using the sites www.tunecaster.com (an online music encyclopedia), www.rockmusictimeline.com (an online chronology of rock lyrics and rock history), www.boomerslife.org (a site with music, people and issues connected to a certain era) but most prominently www.digitaldreamdoor.com (list “created by knowledgeable people and visitors dedicated to the best possible ranking of songs, albums, musicians, and musical artists” (digitaldreamdoor.com)) the top 60 songs of each decade are established and chosen. As mentioned earlier also Wikipedia was used to rule out borderline cases. The reason for the number of songs is to avoid any idiosyncratic features that might occur when using too small an amount of material. After the songs had been decided the

² John R. Taylor describes prototypes in two ways: either the term can be applied “to the central member, or perhaps to the cluster of central members, of a category [...] Alternatively, the prototype can be understood as a schematic representation of the conceptual core of a category.” (1995: 59)

lyrics of each song was collected, using the sites *Lyricsfreak.com*, *sing365.com*, *azlyrics.com*, *stlyrics.com* and *lyrics007.com*. Each lyric was then processed in the text editor Notepad and saved as a simple text document with artist/band, title of the song and the year it was released as its filename: [Elvis Presley – Hound Dog – 1956]. The next step was then to investigate these files by means of a tool or program, such as AntConc. AntConc is a freeware corpus analysis tool (downloadable at <http://www.antlab.sci.waseda.ac.jp/software.html>). In this case a lemma list (downloaded from AntConc) was also used in the research, together with a stop word list (downloaded from <http://www.softexe.com/askw-stop-words-list.html>). These were both used to exclude words that would not contribute to the research.

When working through the collected material, the results for the total amount of text were investigated for a synchronic research as well as diachronic research. That is, the overall word usage was investigated as well as the differences between decades. This was then to be compared to previous studies in this subject.

The results of this quantitative research are then the subject of more qualitative research, investigating what explanations there are to these findings.

3.2 Difficulties in compiling a corpus

A main issue when compiling a corpus is not the amount of material found, but rather the validity of the material. As mentioned earlier, which songs count as rock and which do not? The question is then as follows: which of all the existing rock songs are suitable to include into a corpus? In the gathering of material for this research, repetitions had to be excluded in order to achieve as accurate a result as possible. If this had not been done, song lyrics with a high number of repetitions would have had too high an influence on the outcome. Repetitions have therefore been replaced with ellipsis. An illustration of this is the song *Tutti Frutti*, by Little Richard, where repetitions have been replaced with ellipsis (originally 51 lines long):

- (1) Bop bopa-a-lu a whop bam boo
Tutti frutti, oh Rudy
...
... bop-a-lu ...
Got a girl named Sue, she knows just what to do
...
She rock to the east, she rocks to the west
But she's the girl that I know best
...
... Daisy, she almost drives me crazy
...

She knows how to love me, yes indeed
Boy, I don't know what you're doin' to me ...

When compiling this certain corpus, one has to find suitable and trustworthy sources for lyrics. As mentioned above the sites *Lyricsfreak.com*, *sing365.com*, *azlyrics.com*, *stlyrics.com* and *lyrics007.com* have proven to be the most reliable sources of lyrics. Nevertheless, the words might differ slightly from site to site, which consequently raises the question: which version is the correct one? However important as this may seem this has not been taken into consideration since this small margin of error would not affect the outcome.

Compiling a corpus is time-consuming and at times frustrating. Therefore, definitions are of great importance. In this particular research Wikipedia has worked as a useful key.

4. Results

In this section, the results of the corpus research will be presented together with comparisons with previous studies.

300 lyrics in total were collected from the 1950s onwards to the 1990s (i.e. 60 from each decade), generating in 52,907 words. In order to answer the question which words are the most frequent in rock music lyrics, all the lyrics were processed in AntConc. In the following table, the fifty most frequent word forms figuring in rock lyrics have been collected (also contracted forms and parts of contractions are included in this table):

Table 4.0 The 50 most common word forms

Rank	Freq.	Word
1	2394	i
2	1929	you
3	1917	the
4	1240	and
5	1199	a
6	1089	to
7	849	it
8	783	s ³
9	781	me
10	737	t
11	691	my
12	666	in
13	532	that
14	497	of
15	451	on
16	438	all
17	429	your
18	422	be
19	407	can
20	386	m
21	381	we
22	345	don
23	344	for
24	331	but
25	316	she
26	310	just
27	301	no
28	298	oh
29	290	with
30	286	know
31	275	got
32	272	ll
33	265	is
34	265	love
35	258	so
36	247	was
37	245	when
38	243	baby
39	242	like
40	230	get
41	224	now
42	218	re
43	216	do
44	211	yeah
45	207	go
46	206	what
47	202	this
48	201	they
49	195	ve
50	188	if

^{3 3} Important to point out is that *s*, *t*, *m*, *don*, *ll*, *re* and *ve* are not independent words, but parts of lemmas or contractions of words. One might suggest that these should have been excluded from the table. However, their occurrence does not greatly affect the results. This is also applied in table 4.1.

To be able to generate only lemmas, a lemma list (provided by AntConc) was imported into AntConc. No hyphenation is included in this search and all data is treated as lowercase. The result is shown in the table below:

Table 4.1 The 50 most common lemmas

Rank	Freq.	Word
1	3866	i
2	3169	you
3	1917	the
4	1736	be
5	1277	a
6	1240	and
7	1089	to
8	849	it
9	737	t ⁴
10	666	in
11	565	that
12	534	get
13	497	of
14	451	on
15	448	have
16	438	all
17	410	will
18	407	can
19	381	we
20	345	don ⁵
21	344	for
22	343	know
23	331	but
24	317	go
25	316	she
26	310	just
27	301	no
28	298	oh
29	297	love
30	290	with
31	277	do
32	258	so
33	252	say
34	245	when
35	243	baby
36	243	like
37	239	this
38	224	now
39	218	re ⁶
40	213	come
41	211	yeah
42	207	make
43	206	what
44	205	see
45	201	they
46	194	time
47	188	if
48	181	down
49	181	up
50	179	there

⁴ Part of the contraction *n't (can't etc)

⁵ Part of the contraction don't

⁶ Part of the contraction *'re

Table 4.0 and 4.1 show a high occurrence of the pronouns *I* and *you*. Biber (1991) refers to involvement as “those linguistic features which reflect the fact that speaker and listener typically interact with one another, while writer and reader typically do not.” (1991:43) Biber argues that speakers use, among others, second personal pronouns (*you, your* etc.) due to this interaction. First person pronouns (*I, me, we* etc.) are used when concerned with the expressions of their own thoughts and feelings. These, according to Biber, are some of the characteristics of *involved* text. Biber also talks about detachment (the opposite of involvement) and how it refers to “the characteristics of typical writing which results from the fact that writer and reader usually do not interact” (1991:73). Passives are one of these characteristics and these are also found in the ROLC, one example being Prince's “Let’s go crazy”

- (2) Dearly beloved
We are gathered here today
To get through this thing called “life”

To clarify then; spoken language is more involved than written language, which consequently is more detached. The issue whether rock lyrics are more spoken-like or written-like is brought up further in section 4.4, in this essay. In the mentioned section, the twenty most frequently used words in the spoken and fiction subcorpora in COCA are presented. This was done in order to come to a conclusion whether rock lyrics are more spoken or written-like.

4.1 Comparing genres

Since a majority of the reference material found for this paper concern the general emotion of the lyrics, it is difficult to compare the concrete word usage and frequency between genres. However, in Mukherjee and Kreyer's pilot study of the Giessen–Bonn Corpus of Popular Music (GBoP), they rank the 23 most frequent lemmas in pop song lyrics. The result in this particular research (table 4.1) is very similar to their results (table 4.2); a majority of the 23 most frequent lemmas found in pop music (in the GBoP) is also found among the 23 most frequent lemmas in rock music. The exceptions are *I'm* (since both the headwords *I* and *be* occur in this corpus *I'm* is not included in the ROLC), *for* (appears in place 21), *so* (appears in

place 32), *like* (appears in place 36), *don't* (however, *t* appears in place 9 and *don* appears in place 20) and *love*, which is found in place 29.

Table 4.2 The 23 most frequent lemmas in GBoP

N	GBoP				Word	Freq.	%
	Word	Freq.	%				
1	YOU	6.818 ⁷	3.87	12	YOUR	1.592	0.90
2	I	6.662	3.78	13	I'M	1.568	0.89
3	THE	5.607	3.18	14	ON	1.502	0.85
4	TO	3.912	2.22	15	OF	1.439	0.82
5	AND	3.863	2.19	16	ALL	1.304	0.74
6	ME ⁸	3.745	2.12	17	FOR	1.212	0.69
7	A	3.542	2.01	18	SO	1.206	0.68
8	IT	2.514	1.43	19	BE	1.196	0.68
9	MY	2.214	1.26	20	LOVE	1.142	0.65
10	IN	2.002	1.14	21	LIKE	1.117	0.63
11	THAT	1.806	1.02	22	WE	1.082	0.61
				23	DON'T	1.078	0.61

One could then come to the conclusion that it does not differ that much, fundamentally, between the two genres. However, the result in table 4.1 consists chiefly of Stop Words, i.e. terms that appear so frequently that they lose their usefulness as search terms. A list of stop words was therefore inserted into AntConc, in order to exclude these from the search. Non-content words like conjunctions, determiners and prepositions (often called function words) were excluded, and the results are shown in table 4.3:

Table 4.3 The 20 most frequent lemmas in ROLC (no stop words included)

Rank	Freq.	Word	11	140	gonna ⁹
1	303	know	12	126	feel
2	298	oh	13	119	night
3	297	love	14	118	man
4	275	get	15	99	heart
5	243	baby	16	97	life
6	243	like	17	91	girl
7	211	yeah	18	89	eye
8	194	time	19	84	wanna

⁷ Standardized type-token ratio

⁸ Both *me* and *my* were excluded from table 4.1 since they were considered to be part of the lemma *I*.

⁹ Both *gonna* and *wanna* could be classified as stop words. Due to their research value they have been included in this table.

9	144	way	20	83	give
10	141	day			

Table 4.3 shows that the single most frequent non-stop word in the research was *know*, followed by *oh* and *love*. In Logan et al.'s (2004) analysis of song lyrics of different genres, their results within the rock genre to some extent agree with the findings in this research:

Table 4.4 The ten most frequent non-stop words for selected genres (based on Logan et al.'s analysis)

Reggae	Country	Newage	Rap	Rock
girl	love	adis ¹⁰	I'm	I'm
lover	I'm	go	like	love
know	just	say	get	don't
love	don't	day	got	know
I'm	know	night	don't	just
let's	like	love	n****	like
mi	got	sky	know	got
shout	time	says	s***	you're
like	heart	ergo	ain't	time
gal	go	heart	yo	oh

However, since all stop words have been excluded from this particular research (and Logan et al. apparently not having done so, considering that words like *I'm*, *you're* and *don't* are present in their result), the outcomes obviously are not identical. Words like *love*, *know*, *like*, *time*, *got/get* and *oh* all figure in the top twenty of the collected rock lyrics, and these are also present in Logan et al.'s result. As shown in table 4.4, country is also very similar to rock. Looking at the mentioned table, every genre (except country) has certain genre-specific words: reggae: *girl*, *shout*; newage: *sky*, *day*, *ergo*; rap: *get*, *yo*; and rock: *just*, *oh*.

It is shown that certain lyric-specific words occur in each genre. However, why these certain words were frequent were not stated, and is difficult to explain without further investigation. Certainly, each genre has a certain kind of listeners, with a certain kind of background. Nevertheless, why they use a certain set of words can only be speculated.

One important discovery concerned the fact that several words, which occur in rock music, do not occur in pop music. The words *have*, *can*, *will*, *know* and *get* occur in rock

¹⁰ The proper meaning of this word has not been found

music, but not in pop music, and it seems as if rock music in general is more action-oriented than pop music, in the way it uses numerous verbs.

Moreover, in Katznelson et al.'s study the top ten nouns by genre were listed (table 1.0). All genres showed a deficit of nouns in the top one hundred words, and this is also the case in the current essay: only nine words out of the top hundred were nouns (*love, baby, time, way, day, night, man, life, heart* and *eyes*). However, when comparing the top ten words from ROLC with Katznelson et al.'s study, ROLC has in fact more in common with the nouns listed under country than the ones listed under rock. These differences in results are probably due to the different time spans covered in these corpora (Katznelson et al.'s study only covers the past twenty years).

4.2 Interjections

It is established in the corpus that different utterances expressing emotion or sentiment figures frequently in rock music lyrics. The interjection *oh* ranks as high as second among the most frequently used words while *yeah* positions itself in seventh place. *Ah* is the third most frequent interjection in place 44.

Interjections in song lyrics are difficult to specify since they fulfill different purposes. In Neal R. Norrick's paper (2009:868) he counts that the most frequent interjections (functioning as pragmatic markers) included *oh, ah* and *yeah*; *oh* and *ah* counting as primary interjections whereas *yeah* counted as a secondary interjection. He continues:

“[...] many primary interjections do not express emotions, as is often maintained of interjections generally, but rather information states¹¹. It is usually up to the secondary interjections like *damn, fuck* and *shit*, often with their roots in religion, sex and scatology, to express strong emotions” (2009:868).

With that in mind, the definition that Norrick gives interjections, versus the interjections occurring in rock lyrics, does not seem to match. Looking through the lines where these interjections figure, it seems a great majority of them work as padding. That is, they are used to fill gaps in the lyrics where the melody requires it (or else the rhythm might be disrupted). Examples of this padding can be found in Little Richard's “Tutti Frutti”, as seen in (3) below;

¹¹ Interjections like *oh, huh, hm, ah* are “signaling some change in the speaker's cognitive state.” (2009: 875)

- (3) Bop bopa-a-lu a whop bam boo
 Tutti frutti, *oh* Rudy
 [repetition]
 A whop bop-a-lu a whop bam boo

If it were not for the melody, acting as a template for the words to fit into, the usage of these interjections, or padding, would have been highly excessive. Moving on, looking for occurrences where interjections do not act as padding, we do not find that many instances. However, in Meatloaf's "I'd do anything for love" (4) we find one:

- (4) But I'll never forget the way you feel right now, *oh no*, no way.

The interjection *oh* in this case rather works as an intensifier to the subsequent word *no*, than as padding for the melody. Norrick (2009) talks about this when revealing his finding that interjections function as intensifiers when they, in initial turn position, precede bare *yeah* and *no*, as is the case in (4) above.

4.3 Non-standard spellings

Mukherjee & Kreyer mention different deviant spellings figuring in pop music, for instance *gonna* or *going to*. This spelling, together with *wanna* and *want to*, figures repeatedly in the ROLC, reaching 225 occurrences (shown in diagram 4.5)

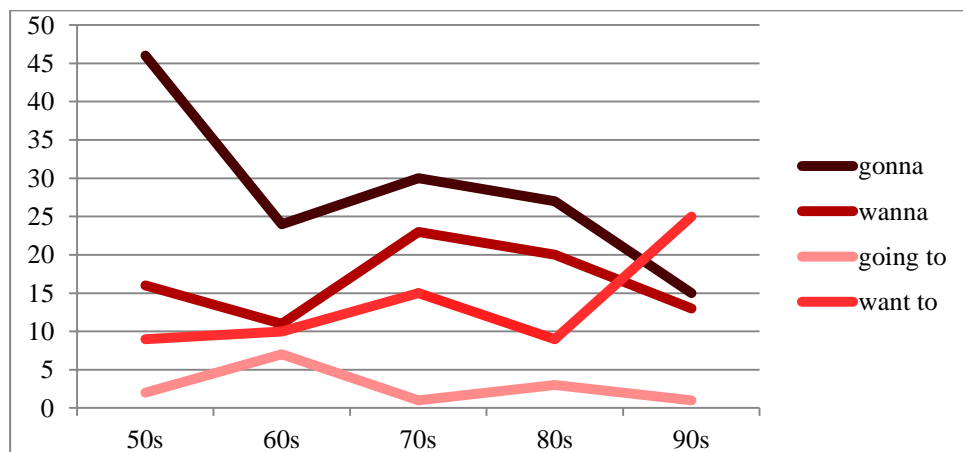


Diagram 4.5 Frequency of contractions, plus their full form, over time

Diagram 4.6 below shows the frequency of *gonna/going to* and *wanna/want to* in COHA (Corpus Of Historical American English).

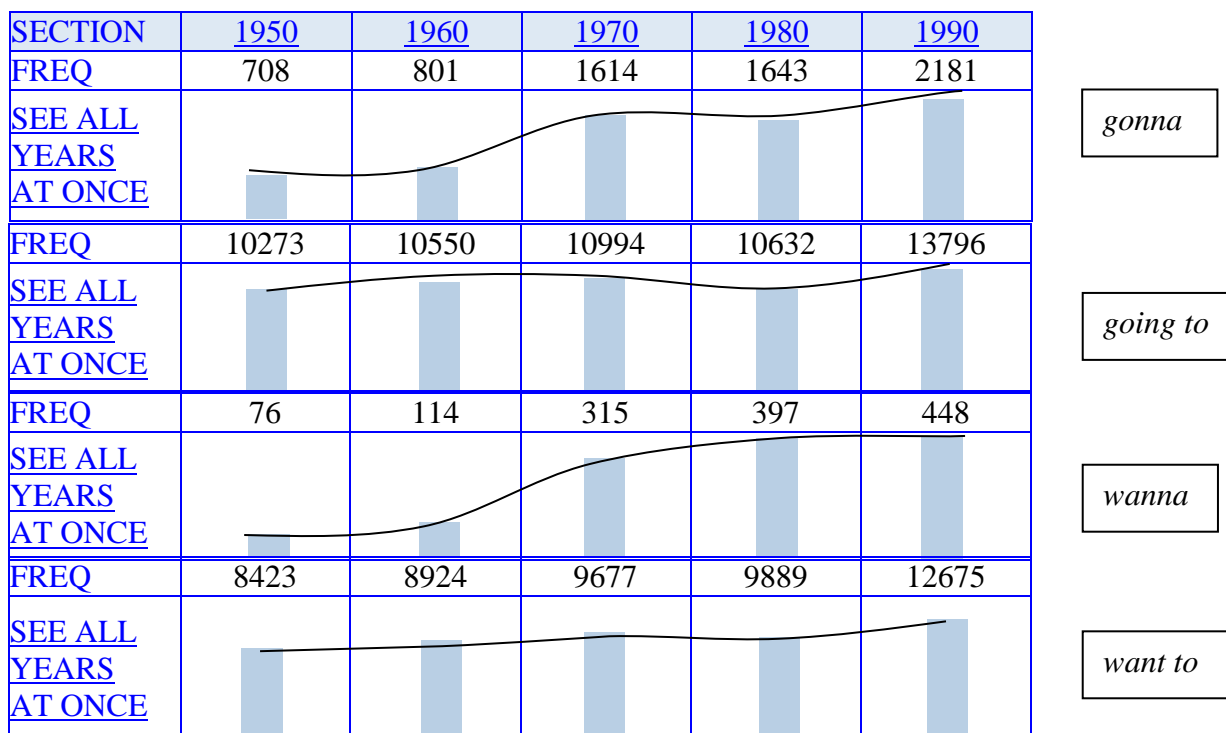


Diagram 4.6 Frequency of *gonna*/*going to* and *wanna*/*want to*, over time, in COHA

As shown in diagrams 4.5 and 4.6, the diachronic frequency curve of *gonna* in rock music does not follow the frequency curve in COCA – they are rather opposites – whereas *wanna* in rock music follows COCA, to some extent, during the 60s and 80s. Moving on, the frequency of *going to* in ROLC does not follow the frequency in COCA, whereas *want to* in ROLC largely follows COCA. To sum up, in COCA, and consequently the overall language, both forms of *gonna* and *wanna* seem to be increasing. Consequently, this indicates that the language used in rock music lyrics does not, in this factor, mirror the general language use in society. This result is surprising, though it is important to highlight the factor that in some instances the pronunciation differs from the spelling. For instance in the song “I don’t want to miss a thing”, by Aerosmith, the question whether the lyrics read “I don’t *want to* close my eyes” or “I don’t *wanna* close my eyes” could be discussed. Due to these kinds of instances, the results in ROLC might be questioned.

4.4 Comparing fields

Moving on, are there any similarities between the word usage in rock music lyrics, and the overall usage in the English language? And consequently, is the word usage in rock music more similar to written than to spoken English? To find out, a corpus called *Genre Frequency*

(http://www.wordfrequency.info/files/genres_sample.xls) was downloaded from COCA (the Corpus of Contemporary American English). In this corpus, the frequency of each of the top 60 000 lemmas – in spoken, fiction, popular magazine, newspapers and academic – is arranged. The spoken and fiction subcorpora were chosen and compared with the word frequency in rock lyrics. The spoken sub corpus consists of 5037 types, whereas the fiction sub corpus consists of 5459 types. It should be borne in mind that the material used from COCA spans over a different time period than ROLC, and of course concern different genres. Since the COCA corpus only uses up-to-date frequency (from the 1990s up to today), and the ROLC is based on lyrics from the 1950s and onwards, the comparison is somewhat unfair. However, through this comparison, we obtain a brief overview of the similarities between rock lyrics and spoken versus written English. The two subcorpora were both loaded into AntConc and the results are shown in table 4.6 below:

Table 4.6 The 20 most frequent words in ROLC and COCA (spoken and written subcorpus)¹²

	ROLC		Fiction		Spoken	
	Word	Freq	Word	Freq	Word	Freq
1	I	2377	the	4092394	the	3859682
2	me	768	I	1382716	I	1346545
3	And	516	she	798208	they	609735
4	You	390	they	352405	would	233766
5	love	289	me	320073	she	212920
6	know	284	would	229865	me	180076
7	got	259	could	207758	one	172827
8	baby	215	one	147159	could	119869
9	like	214	ask	93681	more	90459
10	But	200	head	83140	more	86107
11	time	166	woman	78706	all	78464
12	Oh	161	only	76074	should	77547
13	yeah	144	stand	67559	ask	68414
14	It	141	more	64964	put	68277
15	The	138	put	59445	woman	60731
16	She	136	more	54641	problem	58392
17	well	135	should	47244	oh	55348
18	We	134	place	47236	point	52598
19	gonna	123	pull	46029	only	51930
20	when	122	all	46018	government	51238

¹² In this research additional words were added to both the stop-list and lemma list to match the other sub corpora

COCA's fiction and spoken sub corpora and the ROLC have four words in common (*the, I, me* and *she*), whereas the spoken sub corpus has one unique word in common with the ROLC (*oh*). This is not a big enough difference to claim that rock lyrics are more spoken than written-like. The conclusion thus reads that rock lyrics share features with both written and spoken language and cannot be placed in one category exclusively. However, as mentioned earlier, the high frequency of both first person pronoun *I*, and second person pronoun *you*, suggests that the language used in rock music tends to be more spoken-like than written-like.

Some examples of words that were frequent in ROLC, but not in COCA, are *love, baby, but, gonna* and *yeah*; words that could be considered as parts of an informal language. This then suggest that the language used in rock lyrics is more colloquial than the language figuring in COCA.

4.5 Diachronic results

The material collected for the corpus used in this research stretches over five decades; the 50s, 60s, 70s, 80s and 90s. As mentioned earlier, 60 song lyrics from each decade were selected and collected for this corpus, in total; 300 songs and 52,907 words. To be able to compare the word usage of the five decades, they were all, separately, put into AntConc. The results are displayed in table 4.7 below (no lemma forms or stop words were used in this analysis):

Table 4.7 The most frequently used words during each decade.

	50s	60s	70s	80s	90s
Rank	Word	Word	Word	Word	Word
1	baby	love	Time	oh	know
2	love	yeah	Know	love	love
3	oh	know	oh	know	get
4	rock	get	get	get	like
5	know	baby	like	like	oh
6	get	oh	baby	time	yeah
7	gonna	hey	day	feel	feel
8	like	night	love	life	time
9	roll	ah	man	yeah	life
10	bye	like	run	wanna	day
11	day	girl	yeah	beat	night
12	yeah	gonna	sing	baby	head
13	man	woman	place	lord	change
14	blue	feel	gonna	man	eye
15	crazy	time	boy	gonna	give

16	hear	turn	die	eye	heart
17	fool	man	home	heart	black
18	shake	people	leave	day	baby
19	heart	day	girl	show	dream
20	bop	walk	face	world	long
21	o	glad	wanna	start	care
22	time	heart	call	u	cry
23	true	pretty	give	wait	sun
24	hey	turnin	long	night	world
25	long	give	change	fight	die
26	play	home	dream	leave	light
27	girl	leave	hear	radio	lose
28	tonight	light	walk	tonight	close
29	wanna	lonely	night	call	gonna
30	johnny	lose	put	cry	lonely

A great many of the words frequent in rock lyrics during one decade are also frequent in all the other decades. However, some of the words in table 4.7, which are emphasized and marked by a certain color, change considerably over time. Some of these have been put into diagram 4.8:

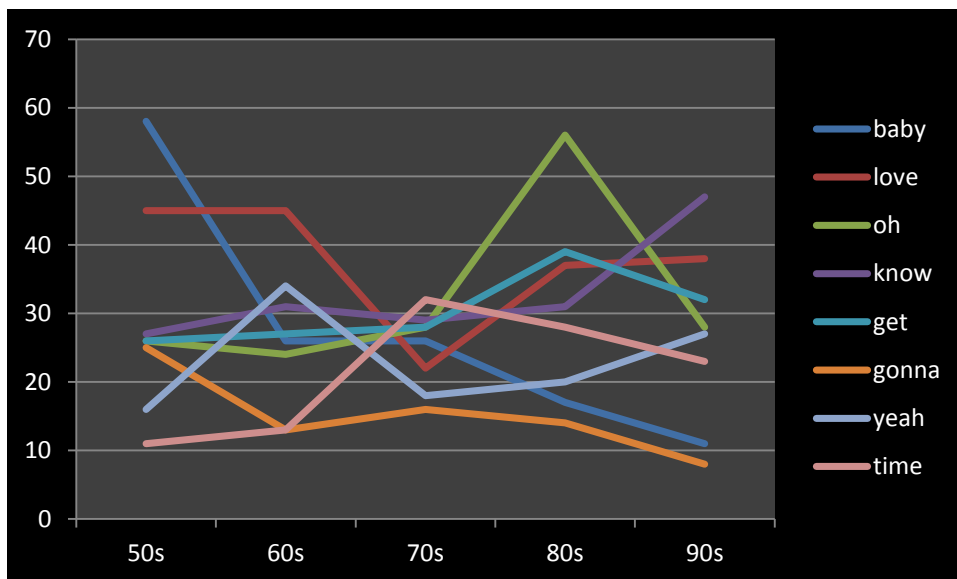


Diagram 4.8 Diachronic changes in frequency for selected words (per 10,000 words)

One word that has had a clear pattern over the decades is the word *baby*. It had a distinct decrease between the 1950s and the 1990s, most obvious between the 50s and 60s. A loose speculation about the reasons for this decrease could be that the word became unpopular in an increasingly more equal society or, simply, that the word quickly became old fashioned and worn out.

Moving on, the interjection *oh* has since the 50s been a highly frequent word, with its lowest rank in place six. As mentioned earlier, *oh* is a popular way of filling out gaps in the melody. Looking through lyrics from each decade the interjection's purpose has not changed:

(5) *Oh*, when you're rocking and rolling

Can hear your mama call

(Little Richard:1958)

(6) You know it's true

Everything I do, *oh*, I do it for you

(Bryan Adams:1991)

Love is one of the most frequently used content word in rock music and is frequent in almost every genre, as shown in table 4.4. The verb form and noun form of *love* engross roughly equal proportions. Kreyer & Mukherjee's investigate how the subject and the object are realized in the GBoP. The figures they produced show that the subject *I* (47 absolute frequency) is the most frequent, followed by object *you* (37 absolute frequency). This is also the case in ROLC, apart from the fact that it does not differ that much between the subject *I* (31) and the object *you* (34). This indicates that in the majority of sentences involving *love*, the subject *I* and object *you* interact in some way (*I love you*). This thus indicates that *love* typically works as a verb, and this realization has provided this essay with its opening quotations. The first one, by Elvis Presley, represents the 50s where approximately 60% of the occurrences of the word *love* belong to the word class verb. The second quotation is by the artist Meatloaf, and it represents the 90s. However, in this case, the results are rather the opposite, since 60% of the occurrences belong to the word class noun.

Moving on, rather than isolating each word, a certain pattern can be spotted in table 4.7. As mentioned earlier, DeWall et al demonstrate that feelings of loneliness and social isolation rose 250% between 1985 and 2004. This may be manifested in the collected material of which ROLC consists. In the 1990s, six of the thirty words included are, all in obvious ways, negative: *night*, *die*, *black*, *cry*, *lonely* and *lose* Comparing this with the only negatively colored word in the 50s, *blue*, the difference is quite obvious. DeWall et al argue that "evidence suggests that individual mental health has decreased substantially over time" (2011:201). The findings in ROLC might indicate that rock lyrics, in this particular matter, do in fact mirror the overall emotion of society, and consequently the language used in it.

In contrast to the former, emotional analysis, the following analysis rather concerns the quantitative perspective of the corpus. Diagram 4.9 shows the total number of words per decade:

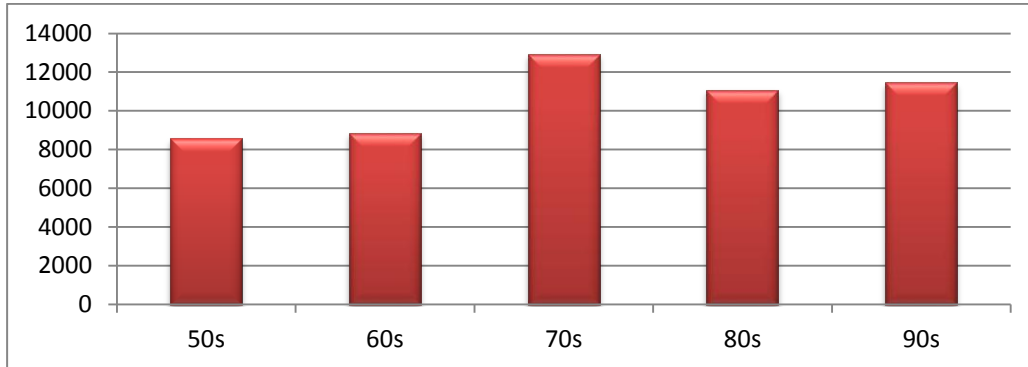


Diagram 4.9 The total number of words for each decade.

As shown in the diagram, the number of words was significantly higher during the 70s. However, since repetitions have been erased, the comparatively low number of words in the other decades might indicate a high frequency of repetitions. Nevertheless, in Pettijohn & Saccos analysis of popular billboard songs across conditions of social and economic threats , they come to the conclusion that ”when social and economic times were difficult and threatening, songs with more words per sentence and more future references were popular” (2009:301).

5. Conclusion

I argue that research concerning the language used in rock music has only just begun. By attempting to respond to the questions initially asked, the hopes of broadening this subject have been fulfilled, and this conclusion will give a recap of the research questions and their answers.

The initial tables in the result section, showing the word frequency of rock lyrics, quickly answer the question *which words are the most frequent*. They show that *I*, *you* and *the* are the three most frequently used words in rock music, ranging from the 50s to the 90s. These findings somewhat agree with the frequency in COCA, except for *you*, which has a surprisingly high frequency in ROLC. This indicates that rock lyrics have a high level of involvement, compared to COCA. *The linguistic features of rock lyrics* are presented further on, and the conclusion suggests that the language used in rock music tends to be more spoken-like than written-like. In section 4.2, distinctive features are presented in the name of interjections. However, these outbursts rather work as padding than as expressions of emotion or as intensifiers.

Moving on, the question whether *the rock genre is comparable with other genres* was addressed next. The ROLC was compared with the GBoP (Giessen–Bonn Corpus of Popular Music) and several similarities were found. Logan, Kositsky and Moreno's research also helps with shedding new light on the similarities between rock and other genres. Words like *love*, *know*, *like*, *time*, *got/get* and *oh* all figure in the top twenty in ROLC, and these are also present in table Logan et al.'s result. The results show that also country is, in several ways, similar to rock. There are certain lyric-specific words which occur in each genre. However, the reason for why certain words are frequent in certain genres needs further research.

The results showed that rock lyrics, and the overall language usage, did not match in the use of *gonna/going to* and *wanna/want to*. However, in section 4.4, ROLC was compared to both spoken and written language. Some words existed that were frequent in both COCA as well as in ROLC, such as *the*, *I*, *me* and *she*. The interjection *oh* occurred in COCA's spoken sub corpus, as well as in ROLC. Moving on, looking at table 4.7, a pattern can be distinguished which can be described with Nathan DeWall C. et al.'s study (2011), about the emotional change in lyrics over time. Their study, which suggests that feelings of loneliness and social isolation rose 250 per cent between 1985 and 2004, could explain the high number of negative words occurring in the 90s.

So far, the research questions have been answered. However, the last question, even though occupying a large part of the essay, has been troublesome. In table 4.7, the actual change of several different words over time can be viewed, although the reasons for their change have not been explained. This is the part that I encourage others to explore; the reasons *why* certain kinds of words are frequent today, but were not so yesterday. I encourage others to use tools like AntConc (described in more detail in section 3.1) for their linguistic studies.

Important to highlight is the selection of data. The selection of lyrics obviously affected the outcome of the corpus-research. The sixty most popular songs from each decade were chosen for this research. However, if the data collection would have been larger, or if the selection were to be of another kind, the outcome had perhaps changed. At times, the corpus seemed too limited (when words like *Johnny* appeared in the search). Furthermore, if the collection would have been bigger, the outcome could possibly have been clearer. Several results were not clear enough to be established, and it probably would have been easier to come to conclusions had I included more material.

The most difficult task when writing this essay has been to limit myself. Further investigations would have included a more careful examination of the word patterns figuring in the tables, such as antonyms (*day/night, life/die, give/lose* etc.) and the usage of gender words (*baby, girl, man, boy* and *woman*). Also, research in the overall emotion of rock lyrics, by analyzing metaphors and the like, would undoubtedly have resulted in interesting, new discoveries.

This study might interest academics, linguists and historians (and mere mortals) in the way cultural subcategories (music and consequently rock music) reflect, or even affect, society. With more profound research within this area, one might possibly find that lyric-specific features of rock music have found its way into society's general language, and are traceable far back in modern history.

References

- Atkins Sue, Clear Jeremy, Ostler Nicholas (1991) *Corpus Design Criteria*. Oxford University Press, Walton Street, Oxford OX2 6DP
- Biber Douglas (1991) *Variation across speech and writing*. Cambridge university press.
- Cole Richard R. (1971) *Top Songs in the Sixties : A Content Analysis of Popular Lyrics*. American Behavioral Scientist 1971 14: 389
- De Clercq Trevor, Temperley David (2011) *A corpus analysis of rock harmony*. Popular Music (2011) Volume 30/1. Copyright © Cambridge University Press 2011, pp. 47–70
- DeWall C. Nathan, et al. (2011) *Tuning in to Psychological Change: Linguistic Markers of Psychological Traits and Emotions Over Time in Popular U.S. Song Lyrics*. Psychology of Aesthetics, Creativity, and the Arts 2011, Vol. 5, No. 3, 200–207.
- Katznelson Noah, et al (2010) *American Song Lyrics: A Corpus-Based Research Project Featuring Twenty Years in Rock, Pop, Country and Hip-Hop*. San Francisco State University.
- Kreyer, R & Mukherjee, J (2007) *The style of pop song lyrics: A Corpus-linguistic pilot study*. Anglia : Zeitschrift für englische Philologie.
- Logan Beth, Kositsky Andrew, Moreno Pedro (2004) *Semantic Analysis of Song Lyrics*. Hewlett Packard Labs. Cambridge, USA.
- López Rúa Paula (2010): *Making and breaking the rules: lexical creativity in the alternative music scene*. Language Awareness, 19:1, 51–67
- Nesselhauf, Nadja (2011) *Corpus Linguistics: A Practical Introduction*. University of Heidelberg.
- Norrick Neal R. (2009) *Interjections as pragmatic markers*. Journal of Pragmatics 41 (2009) 866–891.

- Petrie Keith J., Pennebaker James W., Sivertsen Borge (2008) *Things We Said Today: A Linguistic Analysis of the Beatles*. *Psychology of Aesthetics, creativity, and the Arts* 2008, Vol. 2, No. 4, 197–202
- Pettijohn, Terry F. & Sacco, Donald F. jr (2009) *The Language of Lyrics : An Analysis of Popular Billboard Songs Across Conditions of Social and Economic Threat*. *Journal of Language and Social Psychology* 2009 28: 297 originally published online 13 May 2009
- Renouf Antoinette (1993) *Sticking to the text: a corpus linguist's view of language*. *Aslib Proceedings*, Vol. 45 Iss: 5 pp. 131 – 136
- Rentfrow Peter J., Gosling Samuel D. (2006) *Message in a Ballad : The Role of Music Preferences in Interpersonal Perception*. 2006 17: 236 *Psychological Science*
- Schneider Edgar W., Miethaner Ulrich (2006) *When I started to using BLUR: Accounting of Unusual Verb Complementation Patterns in an Electronic Corpus of Earlier African American English*. 2006 34: 233 *Journal of English Linguistics*
- Smith Nicholas, Hoffmann Sebastian, Rayson Paul (2008) *Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations*. *Literary and Linguistic Computing*, Vol. 23, No. 2, 2008. Oxford University Press.
- Taylor John R. (1995) *Linguistic Categorization*. Clarendon Press. Oxford.