

A Corpus Analysis of English Pop Songs

Brian Cullen

This paper presents a brief corpus analysis of the lyrics of recent English-language pop songs to show the current state of the genre. Using word frequency analysis, it examines the type of vocabulary used and the resulting style and content. It also identifies the primary grammatical structures and the norms of the genre.

1. Earlier Analyses of Pop Songs

Despite the popularity of English language pop songs among teenage and young adult learners of English, there has been little attempt made to examine the nature of the language that is used in these songs. One early attempt (Murphey, 1989) analyzed the top 50 songs in English from a 1987 music chart using a word-frequency count and demonstrated that for the purposes of EFL, pop songs were found to offer “short, affective, simple, native texts with a lot of familiar vocabulary recycled, yet vague.” While useful in its time, this study focused primarily on frequency of lexis in the songs and was unable to make use of the powerful concordancing tools available today. In addition, Murphey’s study was carried out in 1987, and 22 years can see a lot of changes in music and lyrics.

While there appear to be few other published accounts of analyses of pop song corpus, analyses for other purposes include stylistic analyses of war lyrics (e.g. Pennarola, 2004), the use of lyrical contents to digitally index songs (Logan, 2004), and the use of basic natural language processing tools for the analysis of music lyrics (e.g. Martinez et al., 2005). However, none of these or related publications look at song lyrics from the

perspective of language acquisition or use by L2 users.

2. The Corpus

For the purposes of the current study, a corpus was compiled of the complete lyrics for the top fifty songs in a selected week on the Billboard charts (August 5, 2007), the most widely accepted pop chart in the music industry. To analyse word frequency and collocations in the lyrics, the application *Concorder Pro* was used. This is an easy-to-use concordancer designed for learners and researchers. The documentation describes a collocation as “An alphabetical verbal index showing the places in the text ... where each principal word may be found, with its immediate context in each place.” As recommended by Kennedy (1998), *Concorder Pro* formats and displays all the occurrences or tokens of a particular type in a corpus and so enables quick and easy study of collocations. The program also returns numerical information about the tokens of each type present in the file such as the number of occurrences and the relative frequency of the words.

3. Word Frequency

The concordancer program gives useful information about the frequency of the words in the songs. Table 1 shows the most frequent words and Table 2 shows some other interesting findings. For comparison purposes, a newspaper corpus of similar length was compiled from recent articles in the New York Times. In the discussion below, the frequency of occurrence for each item has been standardized to a corpus size of 10,000 words for both the newspaper and the song corpus.

Table 1 - Most Frequent Words

	Songs	Newspaper
1	I	a

2	you	of
3	the	to
4	me	and
5	it	in
6	to	that
7	and	said
8	my	is
9	that	on
10	in	for
11	on	with
12	we	he
13	the	it

Table 2

	Song Corpus	Newspaper
Total words	21,377	23,978
Unique words	2,231	5,174
Type-token rate	89.6%	78.4%
Average number of words per line/sentence	5.3 (per line)	18.2 (per sentence)
Average word length	3.82	4.78

In the song corpus, there are 2,231 unique words out of a total of 21,377. So the ratio of unique to total words can be determined as $1-2231/21377$ or 89.6%. This is known as the type-token rate, or concentration rate. The higher figure for song lyrics quantitatively shows the high level of repetition. We can calculate the average number of words per line to be 5.3, obviously much shorter than other kinds of text. We can calculate the average word length as 3.82 characters. Even without carrying out a fuller analysis, we can conclude that modern popular pop songs generally use repetitive short lines with short words in them. In contrast, the newspaper texts use much longer sentences, longer words and much less repetition of

words within the text.

Songs are primarily interpersonal

Pop songs feature a very high proportion of 'I' (459) and 'you' (417) compared to most texts (38.8 and 20.0 respectively in the newspaper text). This agreement with Murphey shows a consistency between the songs of 1989 and 2007 that is not reflected in other aspects. Halliday (1985) has analysed texts in terms of three metafunctions: interpersonal, textual, and ideational. In Hallidayan terms, the text can be seen as having a very strong tenor component and the text is primarily within the field of interpersonal meaning. In other words, the song text is concerned with the relationship between the speaker and hearer. The high frequency of other pronouns reinforces this belief: "me" (242.6) "I'm" (85.9), "we" (69.1), "your" (65.5), "she" (55.6), "he" (38.6), and "her" (30.5). The most frequent pronoun appearing in the newspaper corpus is "he" (70.8), almost always collocating with "said" or "showed", thus suggesting the ideational metafunction of the text and its role in representing the world. To summarize very simply, songs are about "me and you"; newspapers are about "him out there in the world."

Slang

There is a large amount of slang used in these songs, some such as 'shawty', 'rozo', and 'doughty' being unfamiliar to me. In addition to the obvious slang, there are a lot of deliberate misspellings of words like 'damnnn' to mean 'damn' or 'Kristen Dior' to mean 'Christian Dior', although these variations are not always evident in the sound recording. Some abbreviated forms appear in the songs such as "cause", "gonna", "yeah", and "wanna".

Sexual references

The language used in the rap songs is pretty extreme in places. There are a lot of explicit sexual references such as "I taught her how to talk to

me while she take pipe & opened her up & showed her what a real nigga like”, and references to violence such as “Cause after i beat ya baby i’m liable to fuck up ya whole life”. Indeed, there is little innocence in these songs. Men are tough. Everyone is cool. No-one is coy. In this world, you say and take what you want.

Words and Themes

When I looked at the content of the lyrics, 76% of the songs were clearly about love or relationships in its various forms: breaking up (9 songs), missing someone after a relationship has ended (4 songs), hunting for a partner (4 songs), strongly sexual (4 songs), “there’s a problem, but we’ll be ok” (4 songs), marriage (2 songs), apologizing (1 song) and falling in love (1 song). The only other common theme in the songs was boasting, usually occurring in the songs with gangster-related lyrics. The remaining few songs had miscellaneous themes such as dancing, music, and following a dream.

These songs are dominated by themes of love, so in addition to the high frequency of pronouns, it is not surprising to find that the most common nouns include: “girl” (47.5), “love” (47.1), “baby” (35.4), “man” (24.2), “night” (21.1), “home” (18.4), and “time” (24.2). The most common verbs are generally straightforward indications of action, possession, or desire: “do” (68.1), “got” (45.8), “want” (28.7), “say” (26.5), “come” (24.7), “wanna” (34.1), “gonna” (24.2), “think” (24.7), “give” (23.8), “can’t” (23.3), “show” (15.7), and “look” (9.0). The essence of these songs is that someone wants something and might just act to get it.

Except for a handful of words such as “whine”, “crank”, “hoe” and “rockstar” (all of which have higher frequency than would be expected because of intense repetition), the most frequent 300 words (apart from the slang words) that appear in the corpus are all within the vocabulary of a Japanese first year senior high school student. While usage of words may sometimes differ in lyrics such as “You’re so delicious”, the vocabulary level of an average English pop song should not pose much

difficulty for the average Japanese high school student.

There is just one song in the corpus in which the meaning is not clear. In general, Japanese conversation is less direct than English. Similarly, Japanese songs tend to be more abstract and less direct than English. This built-in ambiguity may not be perceived as a problem by the listener since songs may of course be open to more than one interpretation. However, in the corpus of English pop songs, there is only one song in which the meaning is not clear, indicating that pop songs at least tend to be very direct and clear in meaning. Successful pop song lyrics generally have strong images, a clear story, and most importantly a powerful communicative impact.

Question words

Japanese songs often feature a question aimed at the world in general. Initially, this seemed to be reflected in the English song corpus as there appeared to be a high percentage of question words in the English pop songs, most prominently: “what” (130), “how” (83), “who” (35), and “why” (42). But when we examine the lyrics more closely, we see that these words are not acting as question words, for example: ‘That’s *why* it’ll never work’, “*When* they say it’s over”, and ‘Oh it’s *what* you do to me’. The question words are acting as relative pronouns or conjunctions.

Where true questions appear, some are a type of boast: “whats the chance of you rollin wit me (sic)” and “but why have the cake if it ain’t got the sweet frosting?”. But almost all of the real questions occur in the songs with the theme of “I miss you”. Sometimes, the songwriter is asking questions directly to the partner as to why he/she left. For example, in *Wait For You*, Yamin asks questions to his ex-partner such as the following:

Why did you turn away?

So why does your pride make you run and hide?

Are you that afraid of me?

In other songs, the questions are self-directed. For example in, *When I See*

U, we hear:

What's happenin to me?

In the dark can you tell me what it means?

It is clear that questions of this type are most likely to appear in songs where the songwriter is trying to work out her own thoughts about a situation, what we called a therapeutical or pensieve effect in chapter 1.

4. Norms of the LI Songwriting Genre

Rhyme

One of the primary features of English song that distinguishes it from other languages is rhyme. This is due to the influence of English poetry which imported rhyme from Italian and French poetry in the time of Chaucer. As far back as the 1660's, Milton wrote *Paradise Lost* without using what he termed the "troublesome and modern bondage of rhyming...the Invention of a barbarous age, to set off wretched matter and lame Meeter" (Sansom 1994). However, despite a long history of complaints against rhyme, it is still the overwhelming tendency in English songs, and 80% of the songs in this corpus rhyme in at least some sections of the song. Phrases that rhyme can be easier to remember, reinforce the contours of a melody, help the listener to guess which word is coming next and thus to understand the message, and of course allow humour and word play to be introduced. But the high frequency is probably best attributed to the century-old convention of English song to use them and the corresponding expectation of the listener.

Probably as a result of their more improvised nature, the rap songs tend to feature more near-rhyme (flossing/frosting), (tell/mail). These rap songs also use weak rhyme in the form of assonance such as night/life; pipe/like. In contrast, the adult contemporary songs have many more perfect rhymes (neglect/disrespect).

Song structure

Song structure is not identical across cultures and the English standard verse-bridge-chorus structures do not always correspond exactly to standard Japanese A-melody, B-melody, and Sabi. While all English songs certainly do not follow the same format, there are several standard forms such as AAA, AABA, and verse-chorus which dominate English songs and most L1 songwriters are aware of them, at least subconsciously. Table 4 shows an analysis of song structure of the first twenty songs in the corpus.

Table 2 Song Structure of Corpus songs

- | |
|--|
| <ol style="list-style-type: none">1. ch - v1 -Rap - refrain - ch - v2 - refrain - ch - v3 - ch2. v1 - v2 - ch - v3 - v4 - bridge - v5 - ch3. v1 - ch - v2 - ch - bridge - ch4. Rap - v1 - ch - v2 - ch - v3 - ch - ch5. Bridge 1 - v1 - ch - v2 - Bridge 2 - ch - v36. intro - v1 - hook - ch1 - v2 - hook - ch1 - v3 - hook - ch7. ch - v1 - v2 - ch - v3 - ch - v4 - ch8. intro - v1 - ch - v2 - ch - v3 - ch - v4 - ch9. ch - v1 - ch - v2 - v3 - ch - v4 - ch10. intro - ch - v1 - ch - v2 - ch - v3 - ch11. intro - verse 1 to 6 - ch - verse 7 to 12 - ch - verse 13 to 16 - ch12. ch - v1 - v2 - ch - v3 - ch13. v1 - ch - v2 - ch - bridge - ch14. v1 - bridge - ch - v2 - bridge - ch - bridge - ch15. ch - ch - long rap - ch - rap - ch - rap16. v1 - v2 - ch - v3 - v4 - ch - v5 - v6 - ch17. v1 - ch - v2 - ch - v3 - ch18. intro - ch - rap - v1 - ch - v2 - ch - intro19. v1 - ch - v2 - ch - bridge - ch20. v1 - ch - v2 - ch - bridge - ch |
|--|

In this analysis, it is sometimes not straightforward to define a song section as a bridge or alternative melody. However, it is immediately clear that verse and chorus are still the primary song components. Comparing the structure of songs 9, 11 and 17 shows us that while the number of

verses between choruses may vary, the chorus is still the most important part of the song as it has been in English songs since the time of Chaucer.

Although there are many minor variations, especially at the beginning and end of songs, table 3 shows that most songs follow the relatively standard pattern of alternating verse and chorus. Sometimes the song begins with a chorus. In some songs such as 6, there are two types of repeated chorus, one being labelled as 'hook'. Most of the songs follow v1-ch rather than the more traditional v1-v2-ch. This emphasis on getting to the chorus quickly may reflect shorter attention spans of modern listeners.

Two classic song forms are AAA and AABA. The AAA form (verse 1, verse2, verse 3) does not appear at all in the top 50 pop songs and seems to be losing its significance in the mainstream. The AABA form (verse 1, verse 2, chorus, verse 3) does not generally appear in its original form, but survives with a repeated chorus at the end in the adapted form of v1-v2-ch-v3-ch. The standard for the sixties and beyond, what Weller (2007, p. 48) describes as "stuff like the Beatles and Motown: songs with a verse, a bridge that leads into the chorus, a little middle eight just to change things up, then maybe a little solo or a key change" is still true today. Despite the radical language and explicit reference used in rap, the fundamental song structure remains the same. The primary lesson that can be drawn for songwriters of all popular genres is that the chorus is of supreme importance. Weakness in language could probably be tolerated elsewhere, but probably not in the chorus.

Title

In English pop songs, it is common to find the title of the song appearing as the first line or last line of the chorus. In Japanese pop songs, it is quite common for the title not to appear at all in the lyrics of the song. While this does indeed occur for English songs, in the corpus of 50 songs, all but one of the songs does feature the title within the lyrics, generally in the standard position at the beginning or end of the chorus. In this way, it is easy for the title to become the most memorable line of the song and thus

to help the listener remember the name of the song.

Grammar of English Song Lyrics

Below I have identified some of the common sentence and discourse grammar features of English pop songs.

Spreading sentences over two lines

In English songs, a lyric is often broken over two lines.

Example:

They say we're too young
To get ourselves sprung

Clear subject

Unlike Japanese, English usually employs a clear subject in each sentence.

You're datin' other guys
You're tellin' me lies
Oh I can't believe
What I'm seein' with my eyes

Grammar of Conversation

The use of “like” in lyrics such as “Because I’m in the streets *like* everyday”, particularly evident in the rap songs, is closer to contemporary conversational English than written English. Carter and McCarthy (1997) have shown how the grammar of conversation differs substantially from the grammar of written text. Sometimes the singer may be simply reproducing the “grammatical mistakes” of the English that he hears around him. This is especially true of the rap songs where identity with the street is deemed important (Keyes, 2004).

Non-Standard Grammar

Some of the lyrics follow non-standard grammar which may be derived

from the grammar of conversation (see above) or may be a creative word play by the songwriter.

Examples

1. Imma buy you a drank ooo wee
Ohh imma take you home with me
2. your boy a good look but she my better half
3. I understand that there's some problems
4. But daddy should of never let her out that young

An important question is whether examples like these can be used to justify grammatical errors by Japanese learners of English? The answer must surely be no. Although these examples from lyrics in the corpus deviate from standard written English, they are not mistakes, but rather creative twisting of the English language to match the rhythm of the music. This is qualitatively different from a Japanese user of English who omits the verb of a sentence or makes other basic linguistic errors due to insufficient language ability or carelessness. For example, in the excerpt below from an Japanese songwriter's English, the grammatical and lexical errors make it almost impossible to understand the songwriter's intent.

They look sincere man but they are really contumely
Foolish history repeats itself
People will die that remain know nothing
though people laughing now

When the native speaker deliberately plays with grammar and lexis, it can be creative. When the L2 songwriter does it by accident, it can be interesting, but it was not deliberately interesting and more often than not it will lead to problems in understanding the lyrics.

Problem with analysis

Some shortcomings of this corpus analysis are listed below.

1. The repetition of chorus causes certain words to have a very high frequency. In some cases, this causes aberrations such as the high

frequency of the word “gimme” as a result of the word being repeated 48 times within one section of the song.

2. Rap and hiphop feature prominently on the charts and although the considerable crossover between genres sometimes makes it difficult to specify the genre, 11 of the 50 songs can be classified as rap/hiphop/dance songs. The rap and hiphop songs have many more lyrics than adult contemporary pop songs, presumably because they are said rapidly over a rhythm rather than sung. The rap songs also involve dialogues between two or more of the singers.

However, these figures are conflating the figures from at least two different kinds of songs. The rap/hip-hop songs in the charts are much longer and use quite a different vocabulary to the “adult contemporary” songs. The rap and hiphop songs distort the data because they are much longer. In addition, they have many more repeated phrases.

To get a more accurate picture of language usage, it would be necessary to divide the corpus up by genre into two or more parts. More songs would need to be added for a fuller analysis. Alternatively, separate corpora could be developed from different charts such as adult contemporary. While this would allow more precise analysis, the line between one style of music and another is not always clear, and crossover is occurring all the time. While it is clear that there are differences between genres, it is beyond the scope of this work to analyse them in enough detail to be useful in material design.

References

- Beth Logan, A. K., Pedro Moreno. (2004). *Semantic Analysis of Song Lyrics*.
Bloor, T. & Bloor, M. (1995). *The functional analysis of English : A Hallidayan approach*. London ; New York: Edward Arnold.
Carter, R. & McCarthy, M. (1997). *Exploring Spoken English*. Cambridge University Press.
Condry, I. (2006). *Hip-Hop Japan: Rap and the Paths of Cultural Globalization*. Duke University Press.
Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. Hodder Arnold.

- Kennedy, G. (1998). *An Introduction to Corpus Linguistics (Studies in Language & Linguistics)*. Longman.
- Keyes, C. L. (2004). *Rap Music and Street Consciousness (Music in American Life)*. University of Illinois Press.
- Pennarola, C. (2004). It's a battle of words: A stylistic analysis of antiwar lyrics. *Belgian journal of english language and literatures*, 113-131.
- Sansom, P. (1994). *Writing Poems*. Northumberland: Bloodaxe Books.
- Murphey, T. (1989). *Song and Music in Language Learning*. Peter Lange.
- Weller, P. (2007). All Mod Songs: An Interview with Paul Weller. *UNCUT Magazine*, p. 48.