A Study of Pop Songs based on the *Billboard* Corpus

Yasunori Nishina
Kobe Gakuin University
Japan

Abstract

Listening to pop songs has been without any doubt enjoyed as a pastime all over the world. From the viewpoint of applied linguists, this situation automatically raises the two fundamental questions of what the linguistic features of pop songs are and how pop songs contribute to language learning and education, although they have largely been neglected as a viable source of data or a topic in these fields. For this reason, on the basis of the author's original pop song corpus, this paper investigates various features identified in the lyrics of contemporary popular songs ranked in the Billboard Hot 100 chart for a decade (2002-2011) to grasp the delineation of this genre, and provides basic data utilizable for the design of future English materials and their development in detail.

Keywords: Pop Songs, Corpus Linguistics, Quantitative Analysis

1. Introduction

This paper quantitatively and qualitatively examines trends in modern pop songs and the characteristics of their lyrics, areas in which there has been almost no research to date. Specifically, I have conducted an analysis by compiling a corpus of *Billboard* Hot 100 songs for each of the past 10 years (henceforth, the *Billboard* Corpus) and referencing a variety of attribute information attached to file names.

2. Literature Review

2.1. *Billboard*: calculation of the popularity of pop songs

Billboard continues to have a great impact worldwide on pop songs. According to the explanation of Matsumura (2012), *Billboard*, founded in the nineteenth century, is the largest weekly music industry magazine in the U.S. While it initially contained information on events such as traveling carnivals and theatrical performances, it gradually shifted to music information and is now famous for the *Billboard* Hot 100 chart for popular music, which is an aggregation of such items as retail and internet CD sales, the number of broadcast radio plays, and the number of downloads from cooperating websites. Walker (2016) summarizes the historical changes in the song selection standards for the Hot 100 chart, which are shown in Table 1. In 2005, the song selection method transitioned to the Digital Age System. With such factors as acquisition of data and listening over the internet taken into account, along with purchases of physical CDs, the current selection criteria are more complicated.[1]The corpus for this paper consists of songs ranked on *Billboard*, as it is the most authoritative ranking in America.

Table 1. Historical changes in selection criteria in the *Billboard* Hot100 (excerpted from Walker (2016))

1958-1991	Ranking determined by ratio of singles sales and airplay	
1991	*Billboard* begins collecting sales data digitally (using SoundScan) for quicker and more accurate charts	Analogue Age
1998	*Billboard* drops requirement that song must be released as a single to appear on the chart	
2005	Digital downloads (iTunes) included	
2012	On-demand streaming services (Spotify, Rhapsody) included	Digital Age
2013	Video views (YouTube) included	

2.2. Literature review into pop songs

Walker (2016) and Kreyer(2015) are among the limited research efforts that have quantitatively and qualitatively analyzed the lyrics of pop songs. Walker(2016) quantitatively analyzed year-end Hot 100 songs from 1958 to 2015 using the free statistical software R. Items analyzed included the most frequent words (*love* is the most frequent word), the number of times the artist was ranked in the Hot100 (Madonna had the most with a total of 35, and 1154 artists were only ranked once), the correlation between career history and hit songs in one year (they are in inverse proportion), the diachronic tendencies in the number of words (the average tendency is for overall word count to increase 1.87% a year and for special words to increase 1.36% a year), and the Top 25 particular terms for each decade extracted by a Log-Likelihood (LL: logarithm likelihood ratio) score.[2]

Table 2. Particular terms in top songs in each decade (top 5)

Rank (LL)	1960	1970	1980	1990	2000	2010
1	can dig	boogie	love	pump	wit	imma
2	dig	love	night	cuerpo	club	like
3	oh happy	woman	heart	will	like	bitch
4	miles	doo	shes	ever needed	bum	rack
5	coal	ron	tonight	jam	girl	fuck

Kreyer (2015) also quantitatively and qualitatively studies the relationship between the use of words in pop songs and gender theory by dividing the constructed pop song corpus into sub-corpora of male and female artist groups and conducting such activities as tag analysis classifying the meanings classification of the 30 most frequent nouns and self-descriptive expressions using *I am, I'm, I'ma, Imma* and W-Matrix.

3. Methodology

3.1.Items analyzed

This paper conducts analysis using the original pop song *Billboard* Corpus First, to understand overall trends in *Billboard's* ranked pop songs, I researched (1) the basic data of the *Billboard* Corpus (Tokens, Types, TTR, AWL), (2) the ratios of different genres among the songs, and (3) macro characteristic information other than gender ratio lyrics (vocals). These are effective for understanding trends in sales of popular songs, and they enable knowledge of the characteristics of the current pop song market likely to be popular among such groups as university students, who represent the general public. Concerning word usage in lyrics, I also surveyed the linguistic features of pop song lyrics from a micro perspective through (4) the characteristics of featured songs, (5) suggestions from most frequent words and most frequent N-gram, and (6) qualitative analysis of characteristic patterns.

3.2. *Billboard* Corpus: Basic data

I used *Billboard* Hot 100 Songs from SONGLYRICS know the world (http://www.songlyrics.com/top100.php), a website of hit song lyrics, to build the *Billboard* Corpus. Since this website publishes information on and lyrics fr om the Hot 100 songs for each year from 1950 to 2011, I used the site to gather extractable lyrics on a total of 1,0 00 songs from the past 10 years, and I constructed the *Billboard* Corpus by excluding noise such as leading whites pace through employing regular expressions in CotEditor[3].

Table 3 presents basic information from the Billboard Corpus, i.e., the average total words in each song (Tokens) in the Hot 100 in each year, the average number of different words in each song (Types), their ratio (Type-Token Ratio: TTR), and the average word length (AWL)[4].The average number of Tokens for the 10-year period from 2002 to 2011 was 502, the average number of Types was 149, the average TTR was 30.67, and the average AWL was 3.47.

Table 3. Annual *Billboard* Corpus basic information

Year	Tokens	Types	TTR	AWL	Year	Tokens	Types	TTR	AWL
2002	507	153	30.68	3.49	2007	527	148	29.42	3.44
2003	534	164	31.84	3.49	2008	484	136	28.76	3.46
2004	542	168	31.80	3.49	2009	479	139	29.85	3.48
2005	525	155	30.73	3.43	2010	472	141	30.15	3.50
2006	550	156	29.54	3.42	2011	397	128	33.90	3.52

While Tokens and Types are trending downward, the TTR value itself has not changed much. That is to say, while the economy of lyrics is increasing with each passing year, their nature is maintained quantitatively, and the message they aim to convey (in spite of the content of the lyrics) is becoming more concise.

4. Analysis

4.1. Genre and gender ratio

Table 4 presents ratios by genre of the total of 1,000 songs recorded in the *Billboard* Corpus. In the most recent 10 years, the four genres of HipHop (30.6%), Rock (20.3%), Pop (19.9%), and R&B (16.5%) have been popular. Looking at the shares of other genres, it can be said that they still lack influence on the music scene.

Table 4. Music genre shares for the past 10 years (2002-2011)

Rank	Genre	Count	%	Cumulative%	Rank	Genre	Count	%	Cumulative%
1	HipHop	306	30.6	30.6	9	Blues	5	0.5	98.5
2	Rock	203	20.3	50.9	10	Reggae	4	0.4	98.9
3	Pop	199	19.9	70.8	11	Soul	4	0.4	99.3
4	R&B	165	16.5	87.3	12	Latin	3	0.3	99.6
5	Country	71	7.1	94.4	13	Jazz	2	0.2	99.8
6	Ska	16	1.6	96.0	14	Folk	1	0.1	99.9
7	Electronic	14	1.4	97.4	15	Musical	1	0.1	100.0
8	Others	6	0.6	98.0	Total		1,000	100.0	100.0

Please refer to Table 5 for a summary of the gender composition for the vocals in the total of 1,000 songs in the *Billboard* Corpus. Also, since there have been many featured songs ("F songs") in the recent music scene, Table 5 considered the gender ratio for the main vocals. As a result, we can see that this ratio is roughly 7 to 3, with male musicians predominating.

Table 5. Main vocal gender ratio for the past 10 years (2002-2011)

	Male	Female	Total
Number of songs	684	316	1,000
Share (%)	68.4	31.6	100

Table 6. Comparison of individual artist songs and F songs (total song count)

	F song (N=193(19%))	NF song (N=807(81%))	Significant difference	Test method
Tokens	658	502	**	Shapiro-Wilk
Types	195	149	**	Shapiro-Wilk
TTR	29.93	30.67	*n.s.*	Brunner-Munzel

Note: p*<.05, p**<.01

4.2.Linguistic characteristics of F songs

Next, using the Mac native corpus and concordance CasualConc file search function, I counted F songs from characteristic tags appended to file names recorded in the Corpus. As a result, over this 10-year period there were 193 F songs. Table 6 shows the comparative results of quantitative information for F songs and non-featured songs ("NF songs"). Tokens, Types and TTR are average values for one song.

I selected a quantitative assay method between the two groups based on normality (Shapiro-Wilk test) and homoscedasticity validation results. Since only equal variances were recognized for Tokens and Types and TTR did not allow the Wilcoxon rank sum test for either one, I used the Brunner-Munzel test. As a result, while F songs exceed NF songs for Tokens and Types (with a 1% standard of significance), there was no significant difference for lexical density[5]. Using the results of a diachronic study of *Billboard* ranked songs from 1960 to 2000, Walker (2016) found that total word count was on an upward trend over a 50-plus year period, since compared to the past, the length of recent songs had increased (from 2.5 minutes to around 4 minutes), song tempo had increased, and works by two or more artists had risen in number. One can actually see from Table 6 that F songs tend to have a higher total word count than NF songs, which partially supports the supposition of Walker (2016).

What words are commonly used in F songs? Table 7 compares characteristic words of F songs and NF songs.

Table 7. Comparison of characteristic words in F and NF songs (top 20 words)

Non-Featured Tunes						Featured Tunes					
Rank	Words	LL	Rank	Words	LL	Rank	Words	LL	Rank	Words	LL
1	la	31.93	11	heart	6.85	1	ya	159.03	11	got	59.48
2	re	18.22	12	every	6.07	2	wee	107.02	12	rock	58.51
3	of	16.24	13	everything	6.00	3	wit	75.63	13	double	58.46
4	time	9.89	14	more	5.94	4	runaway	71.04	14	ima	55.24
5	i	9.83	15	turn	5.88	5	lean	69.48	15	like	50.82
6	ve	9.63	16	is	5.70	6	ass	67.63	16	move	47.65
7	goes	9.25	17	were	5.69	7	lolli	61.50	17	niggaz	46.29
8	this	9.17	18	goodbye	5.68	8	ha	60.95	18	ridin	44.31
9	there	7.79	19	oh	5.55	9	suga	59.62	19	skeet	43.86
10	boom	7.33	20	one	5.40	10	get	59.62	20	nigga	40.48

Table 7 shows that expressions that frequently appear in rap music, including slang (*ya, wee, ass, ha, nigga*), are often used as characteristic words in F songs. In particular, there is a clear difference in LL values between F songs and NF songs. *La* (LL = 31.93), which has the number 1 rank for F songs, does not even surpass the number 20 ranked word for NF songs, which is *nigga* (LL=40.48). The expressions in Table 7 can be said to add color to the lyrics of recent F songs.

4.3. What can be said about high frequency words and N-grams?

Table 8 summarizes high-frequency words over the most recent 10 years and high-frequency 3-grams. Rel.Freq. is Relative Frequency and Rel. Ratio is the share in the number of songs among 1,000 songs.

Table 8. High frequency words and high frequency 3-grams (top 15)

Rank	Words	Rel. Freq.	Rel. Ratio	3-grams	Rel. Freq.	Rel. Ratio
1	i	51.47	0.99	i don t	1.87	0.35
2	you	41.08	0.96	i can t	1.26	0.24
3	the	28.63	0.98	la lala	1.13	0.02
4	it	23.20	0.95	nanana	1.06	0.02
5	and	19.39	0.97	oh ohoh	0.99	0.05
6	me	18.90	0.94	don t know	0.92	0.18
7	to	18.59	0.97	i m a	0.87	0.15
8	a	17.50	0.92	iain t	0.86	0.18
9	t	15.32	0.91	you don t	0.85	0.17
10	my	13.55	0.89	and i m	0.65	0.17
11	s	13.25	0.92	ill be	0.55	0.11
12	that	12.00	0.89	don t wanna	0.53	0.09
13	m	11.20	0.81	i m gonna	0.52	0.09
14	in	10.50	0.87	i know you	0.52	0.11
15	on	9.18	0.81	i m not	0.50	0.11

In addition to personal pronouns such as *I*, *you*, and *my*, high-frequency words that are prominent in Table 8 are prepositions such as *in* and *on*. For N-grams, in particular from the results of 3-grams, negative expressions such as *I don't*, *I can't*, *don't know*, *I ain't*, *don't wanna*, and *I'm not* stand out to a very greatextent. Interesting discoveries are shown in order as a result of researching 3-grams in detail.

To begin, 28.66% (133 examples) of the number six ranked negative expression for 3-grams, which is *don't know* (464 examples), have the second-person subject *you*, as in *you don't know*, and of these, first-person-related phrases including *You don't know me* (60 examples), *You don't know my name* (11 examples), and *You don't know who I am* (7 examples) are the majority (about 70%) of objects. That is to say, "second-person + *don't know* + first-person- related phrase" can be understood as a typical semantic sequence within the discourse of Western pop songs (Hunston, 2011; Nishina, 2011).

Next, although the number eight-ranked negative expression *I ain't*is considered an abbreviated colloquial expression of the standard English *am not,*[6] as Fujii (1984) points out, since in African American Vernacular English (hereafter, AAVE) and the Southern U.S. dialect it can even be a substitute for *are not, is not, have not, has not*, *do not, does not, was not, didn't,* and *weren't*, its characteristic is that it is used as "all-person" in that it does not distinguish between person, count or tense. The examples in (1a), (1b) and (1c)below show a few instances of the all-person *ain't* extracted from the *Billboard* Corpus.

(1)   a. Now it's too late I know she <u>ain't</u>comin' back   --- *Burn* by Usher (HipHop)
      b. We <u>ain't</u>gonna hurt nobody   --- *Like That* by Houston (HipHop)
      c. So you <u>ain't</u>got to give my loving away   --- *Hey Daddy* by Usher (HipHop)

Anderson, Carnagey, & Eubanks (2003) report that the lyrics of modern pop songs have the negative effect of increasing aggressiveness in the young generation through continued listening, due to their inclusion of violent and sexual terms and expressions. Actually, of all 431 usage examples of *ain't*, it was accompanied by *I* in the L1 position (in other words, it was a contraction of *am not*) in a mere 126 cases (29.23%); the remaining roughly 70% was recognized as clear non-standard English (N.B. According to an anonymous northern U.S. informant, *ain't*may be considered non-standard even as a contraction of *am not*. Also, no one recognizes a special relationship between *ain't* and *am not* who has not studied the history of the English language).Further, research of the discourse in lyrics using this 70% of *ain't* certainly revealed frequent use of aggressive words (e.g.,*damn, huffing, dope, coke, muthafuckin, niggas, drugs*)(N.B. According to an anonymous American informant, the meaning of "dope" in the subculture was roughly "very fashionable," not something having to do with aggression in some cases. Historically it's also had an "illicit drug" sense.

Also, "niggas" necessarily implies aggression among speakers of AAVE, although it's a derogatory term among white speakers and can be among AAVE speakers as well even when they are talking to other AAVE speakers: it really depends on the context in which it is used). Accordingly, non-standard English that does not adhere to linguistic norms, like AAVE and the Southern U.S. dialect, is popular in modern pop songs, and this trend may be related to the aggressiveness in lyrics that has come to be seen as a problem in recent years. Below is a passage from *Runaway Love*(HipHop)by Ludacris feat. Mary J. Blige. The underlined lyrical content cannot be overlooked in Japanese pop songs. Even if one were to give priority to authenticity, songs with such lyrics probably should not be used in places of education.

(2) Momma's <u>on drugs</u>, gettin' *** up in the kitchen...
    <u>Bringin' home men</u> at different hours of the night...
    She tries to resist but then all he does is <u>beat her</u>...

Furthermore, as shown in example (1b), Labov(1972) found that double (or triple) negative expressions using two (or three) negative terms in one sentence are a peculiar grammatical characteristic of AAVE, with the negative expressions serving not to cancel out the negatives but to emphasize them (N.B. these are now becoming normal negative sentences and the negation isn't especially emphasized, even for some Northern U.S. speakers). In other words, the meaning of the example (1b) is "We aren't going to hurt anyone," with the negative emphasized. Incidentally, a total of 108 examples (25.06% of *I ain't*) this so-called double negative expression of "*I ain't*+ negative word" were found, and the top 3 negative phrases were *no* (40 examples), *nothing/nuttin'* (22 examples) and *never/neva*(21 examples). These were followed by terms such as *nowhere* and *nobody*. One must pay attention to the peculiar spelling in pop songs. Of course, this would likely be difficult to use for educational purposes (N.B. In general, this kind of spelling difference relates more to how much the lyricist wants to use standard spelling and how much that lyricist wants to use a non-standard spelling that in fact better represents the actual pronunciation).

(3)   a. I <u>ain't</u>got <u>no</u> money   --- *The Way I Are* by Timbaland feat. KerlHilson& D.O.E.
      b. I <u>ain't</u>goin'<u>nowhere</u>   --- *Sweet Dreams* by Beyonce Knowles
      c. But I <u>ain't</u>seen <u>nuttin'</u>   --- *Buttons* by The Pussycat Dolls feat. Big Snoop Dogg

4.4. Characteristic preposition expressions

When researching patterns in English, even among function words it is qualitatively speaking most meaningful to extract patterns focused on prepositions (cf. Hunston& Francis, 2000; Francis, Hunston& Manning, 1996; Hunston, Manning& Francis, 1998; Hunston, 2011; Nishina, 2011). In this regard, as this is a pilot study we will look for typical patterns and conduct qualitative research concerning double layering of the prepositions *in* (14th) and *on* (15th) ranked among high frequency words in pop songs in Table 8.

To begin, Table 9 shows the results of gap searches of high frequency 3-grams including *in* or *on*.

Table 9. 3-grams including *in* or *on* (top 10 types)

		in					*on*		
Rank	OR	3-grams	Rel.Freq.	F-Ratio	Rank	OR	3-grams	Rel.Freq.	F-Ratio
1	8	in the *	0.90	0.54	1	29	* on the	0.97	0.38
2	9	* in the	0.90	0.54	2	29	on the *	0.97	0.38
3	78	* in my	0.31	0.28	3	144	on my *	0.45	0.23
4	78	in my *	0.31	0.28	4	145	* on my	0.45	0.23
5	266	* in a	0.16	0.19	5	223	* on me	0.37	0.12
6	266	in a *	0.16	0.19	6	227	on me *	0.37	0.12
7	284	* up in	0.16	0.15	7	421	I * on	0.24	0.12
8	284	up in *	0.16	0.15	8	441	* come on	0.23	0.09
9	287	I * in	0.16	0.18	9	445	come on *	0.23	0.09
10	349	in your *	0.14	0.16	10	537	on a *	0.20	0.13

* OR (Original Rank) shows the original ranking.

Table 9 shows the tendency for extraction of pairs of 3-grams with slots focused around particular pairs of words such as * *in the* * and * *on the**. What also stands out are first and second person possessive pronoun expressions such as * *in my* (third place), *in my* * (fourth place), *in your* * (tenth place), *on my* * (third place) and * *on my* (fourth place). With a focus on the first-person pronoun clusters *I am, I'm, I'ma* and *Imma*, Kreyer (2015) also qualitatively analyzes self-descriptive expressions by gender. This work has high value in actually validating personal-pronoun expressions[7].

Table 10 summarizes the top 10 3-grams including the first- and second-person possessive pronouns *my* and *your*. In all cases, "*in + my/your + X*" is a high-frequency pattern. On the other hand, for *on my* X (Rel.Freq.=0.45), which uses *on*, there is a gap in ranking between *my-your*, and *on your* X (Rel.Freq.=0.18) is unraked in 13th place. Can anything be recognized here, such as not just quantitative differences but, qualitatively speaking, differences among types of stated words in the blank slots? Please also refer to Table 11.

Table 10. 3-grams including *my* or *your* (top 10 types)

|		*my*					*your*		
Rank	OR	3-grams	Rel.Freq.	F-Ratio	Rank	OR	3-grams	Rel.Freq.	F-Ratio
1	78	in my *	0.42	0.28	1	246	your * i	0.23	0.15
2	108	my * i	0.35	0.26	2	341	your * and	0.19	0.16
3	144	on my *	0.30	0.23	3	349	in your *	0.18	0.16
4	187	* my life	0.27	0.13	4	354	* in your	0.18	0.16
5	190	my * and	0.27	0.24	5	392	i * your	0.17	0.13
6	190	my life *	0.27	0.13	6	507	* your love	0.14	0.03
7	311	* my heart	0.20	0.13	7	507	you * your	0.14	0.12
8	311	my heart *	0.20	0.13	8	517	your love *	0.14	0.03
9	381	of my *	0.17	0.13	9	550	* be your	0.13	0.06
10	455	my * my	0.15	0.07	10	550	be your *	0.13	0.06

* The shaded areas are expressions that include prepositions.

**Table 11. High frequency words stated in the blank slots of *in my/your* X (top 10)**

		in one's X				*on one's* X		
Rank	*in my* X	Rel.Freq.	*in your* X	Rel.Freq.	*on my* X	Rel.Freq.	*on your* X	Rel.Freq.
1	head	0.19	eyes	0.10	mind	0.08	mind	0.04
2	life	0.14	heart	0.04	way	0.05	lips	0.02
3	heart	0.06	arms	0.04	tongue	0.05	corner	0.02
4	dreams	0.05	lovin	0.02	baby	0.03	disco	0.02
5	hand	0.05	head	0.02	knees	0.02	own	0.02
6	mind	0.03	ear	0.01	own	0.02	face	0.01
7	arms	0.03	life	0.01	arm	0.02	pretty	0.01
8	eyes	0.03	car	0.01	neck	0.02	side	0.01
9	face	0.03	rock	0.01	chest	0.02	way	0.01
10	air	0.03	closet	0.01	lips	0.02	back	0.01

From Table 11 we can see that body parts in particular (*head, eyes, arms, face, lips, tongue, chest, neck*, etc.) easily arise from the *in/on one's* X pattern. Further, just looking at *in one's* X, *in my* X is preferred for terms related to emotions, memory, and imagination (*heart, dreams, mind*). For reference, Figure 1 shows the gender ratio for vocals using the 10 types of *in my* X high-frequency expressions (absolute numbers), and Figure 2 shows a structural comparison of *in my* X high-frequency expressions by gender.

Figure 1. Gender ratio for in my X high frequency expressions
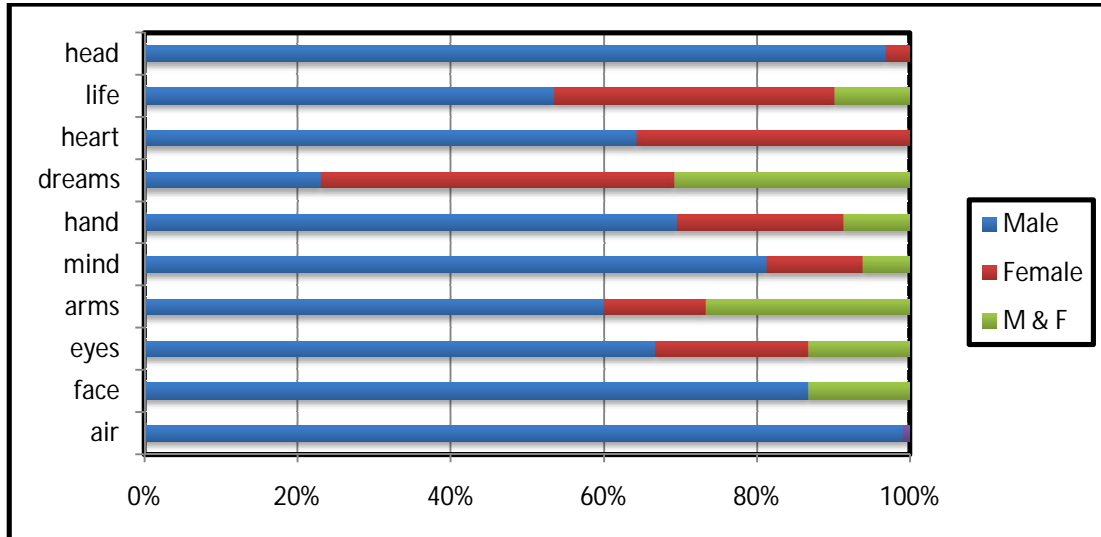
Figure 2. Structural comparison of *in my* X high frequency expressions by gender
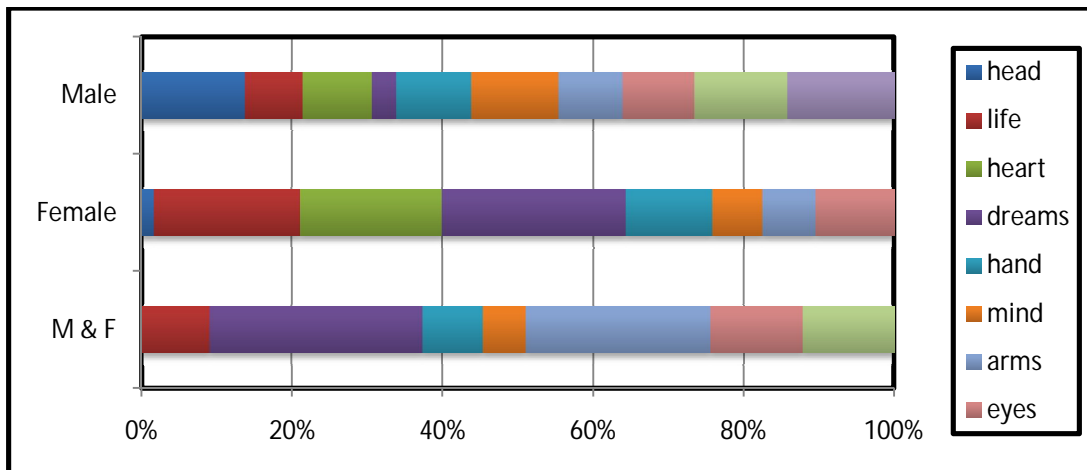
While male vocals were predominant in *Billboard* ranked songs in Table 5, even use of *in my* X was frequent for male artists when stating body parts. However, when making disembodied, abstract statements of dreams, feelings, and life, such as *in my dreams*, *in my heart,* and *in my life*, the use of female vocals stands out.In other words, we see that male songs tend to focus on what are tangible and female songs on the abstract.[8]

We will now focus on qualitative difference between the genders. First, let us refer to *in my arms* as a representative example of a body part expression in which the use of male vocals was predominant. We can see that a man's desire to possess a woman is vividly expressed in song lyrics.

(4) Male desire to possess as seen in*in my arms* (three typical examples)

a. So come over here and lay down <u>in my arms</u> Baby, tell me everything that's on your heart   --- *Love Don't Run* by Steve Holy

b. If I could make it right I'd do it all tonight Hold you <u>in my arms</u>, with you I'd spend my life   --- *Far Away* by Tyga feat. Chris Richardson

c. <u>In my arms</u>, in my mind, all the time I wanna keep you right by my side 'Till I die I'm gonna hold you down   --- *Let Me Hold You* by Bow Wow feat. Omarion

Next, there is a qualitative difference between the genders concerning the use of *in my life*, where the use of female vocals stands out.

(5)Strong thoughts of women as seen in *in my life* (three typical examples for each)

a. But I don't mind as long as I can have you <u>in my life</u>Aaw baby, ''m satisfied, Even if you're not just mine --- *Rock Wit U* by Ashanti

b. Can't you feel my heart beat so, I can't let you go. I want you <u>in my life</u>. --- *Every Time We Touch* by Cascade

c. See I need you <u>in my life</u> for me to stay No, no, no, no, no I know you'll stay No, no, no, no, no don't go away --- *Super Bass* by NickiMinaj

(5a) sings about the strong, painful love of a woman that would be satisfied just to have someone even if they love someone else. (5b) and (5c) describe the desire to not let go of a partner and for a partner who does not leave. Again these expressions use the phrase *in my life*, which acknowledges the importance of a partner.

What we can say from these examples is that there is a tendency for men to prefer physical sensations they can actually feel, a desire for men to possess women, and an overemphasis of one's own body parts. In other words, it can be said that there is a preference for self-asserting expressions. On the other hand, women tend to favor more abstract, disembodied, emotional expressions and overstate emotional connections and their own strong thoughts compared to those of men, who frequently use expressions seeking physical fulfillment.

5. Conclusion

This paper conducted research based on characteristic information of artists and songs as well as words and expressions used in the lyrics of a total of 1,000 songs ranked in the Billboard Hot 100 over the past 10 years, to qualitatively and quantitatively clarify some of the characteristics of modern pop songs. Since very little linguistic analysis of pop songs has been done to date, after conducting multifaceted quantitative and qualitative analysis of the lyrics of pop songs such as that done in this paper I would like to consider such matters as development of effective and efficient teaching materials and methods for studying English. In particular, from the surveys of beginning-level learners conducted by Nishina (2016) (83 first year university students: TOEFL ITP average of 417), we see high expectations for pop songs as learning materials in terms of serving as an effective educational approach from the perspective of motivating students lacking a desire to study English. In the future, I will visualize the selection method for pop songs that should be used in educational materials and classes from the perspective of vocabulary level, expressions used and phonemes and continue to conduct research.

Notes

1. Aside from general rankings that mix these music genres, there are also rankings for all genres, including pop, rock, R&B, hip hop, country, jazz, and ringtones. Incidentally, the ranking criteria in Japan tallies sales, air play, tweets, look ups, YouTube plays, and streaming, and aside from calculation of the Japan Hot 100 there are charts for each genre.
2. In the years between 2000 and 2009, club, go head, and shorty were words and phrases repeated in 50 Cent's "In da Club." Walker (2016) points out that repetition of such specific songs may skew the results of extraction by LL.
3. The most recent Hot 100 from this site is a ranking from 2011. The data I collected is therefore from the 10 years going back from 2011 (2002-2011).
4. While, generally speaking, Tokens means total word count and Types means total different word count, in this paper both mean average word count. Hence, precisely speaking, they are Average Tokens and Average Types.
5. The testing method for the two groups requires selection of an appropriate item after testing for normality and homoscedasticity. For the normality of Tokens, Types and TTR, in Shapiro-Wilk testing the P value of F songs exceeded the significance level (F songs: W=0.99, p>0.05 (Tokens), W=0.99, p>0.05 (Types), W=1.00, p>0.05; NF songs: W=0.95, p<0.01 (Tokens), W=0.88, p<0.01 (Types), W=0.98, p<0.01 (TTR)).For homoscedasticity testing, since I arrived at the conclusion that population variance is equal for Tokens and Types in the two groups (F(192,806)=0.90, p>0.05 (Tokens); F(192,806)=0.98, p>0.05 (Types);

F=(192,806)=0.48, p<0.01 (TTR)), I conducted the Wilcoxon rank sum test, which is a quantitative testing method for cases of homoscedasticity being recognized even without normality. A significant difference was recognized for both as a result (Tokens: W = 121060, p<0.01; Types: W = 121840, p<0.01). Since neither normality nor homoscedasticity were recognized for TTR, the Brunner-Munzel test, which does not need to consider either, did not find a significant difference at a 1% standard (BM=0.62, df=407.73, *n.s.*).

6. Hill (1965) describes *ain't*as historically deriving from the three forms *aren't*, *amn't*, and *haven't*. In AAVE the /d/ in *don't* was frequently dropped, yielding *on't*, and for *didn't*, the second /d/was eliminated to yield *in't*. Since this closely resembled the already extant *ain't*, *ain't* came to be used for *didn't* in AAVE.

7. I gave these the semantic classifications of Independence, Power, Bad boy/Bad girl, Substance abuse, Vulnerable, Romantic, Supportive, and Sex and compared their frequency by gender.

8. *… in my air* is an example extracted from one song. It is an expression taken from *in my air force ones* in the 2003 Nelly song "Air Force Ones."

References

Anderson, C.A., Carnagey, N.L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology, 84(5)*, 960-971.

Francis, G., Hunston, S., & Manning, E. (1996). *Grammar Patterns1: Verbs*. HarperCollins.

Fujii, K. (1984). *Diction in the dialect of the American South as seen in works of literature.* Tokyo: Sanshusha.

Hill, A. (1965). The tainted 'ain't' once more.*College English 26*, 298-303.

Hunston, S. (2011). *Corpus approaches to evaluation: phraseology and evaluative language*. Routledge.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus approach to the lexical grammar of English*. John Benjamins.

Hunston, S., Manning, E., & Francis, G. (1998). *Grammar patterns II: Nouns and adjectives*. HarperCollins.

Kreyer, R. (2015). "Funky fresh dressed to impress": A corpus-linguistic view on gender roles in pop songs. *International Journal of Corpus Linguistics, 20(2),* 174-204.

Labov, W. (1972). *Language in the inner city: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania press.

Matsumura, A. (2012).*Digital Daijisen*. Tokyo: Shogakukan.

Nishina, Y. (2011). Comparison of linked meanings between particular academic disciplines: Direction of future ESAP research. *JACET Kansai Journal, 13*, 14-25.

Nishina, Y. (2016). Building a Theoretical Model for the Usefulness of Pop Culture Teaching Materials. *JACET Kansai Journal, 18*, 52-65.

Walker, K. (2016). 50 years of pop music. Retrieved from http://kaylinwalker.com/50-years-of-pop-music/.