Coffee Break 2: Taylor Swift v.s. Beyonce

Anusha Bhat

November 2023

1   Introduction

Taylor Swift and Beyonce are two of the most prominent and influential female pop artists in the US entertainment industry. Both women have recently released widely acclaimed and well-received full-length studio albums and have embarked on sold-out world tours within the past year. Naturally, these two Grammy-winning singers have often been revered as some of the best artists of this century, resulting in them often being pitted against each other by fans trying to determine which singer is better–but what truly distinguishes their artistry?

In this report, we aim to investigate and elucidate potential similarities and differences between Taylor Swift's and Beyonce's discographies. Specifically, we will explore which words and parts-of-speech are the most frequently used as well as the usage of the over-arching themes of love and hate across each artist's discography. Additionally, we want to identify differences in emotional valence across eacb artist's discography. Although we will not be able to answer the question on who is truly a better artist (if that is even possible), understanding what themes and parts-of-speech are used in well-received lyrics can help clarify stylistic choices made during the writing process and highlight what content is being publicized to impressionable audiences.

We hypothesize that Taylor swift may have more nouns and adjectives than Beyonce since Taylor Swift sings through both first-person and third-person point of view in her songs. However, Beyonce may have more verbs since she employs more colloquial lyrics in her lyrics. Additionally, we hypothesize that both of the artists will have similar prominence of the overarching themes throughout their discographies, however, the context in which these themes are discussed may differ since Beyonce sings about more mature topics than Taylor Swift does.

2   Data

To aide our analysis, we created two corpora of song lyrics with one corpora for each artist's discography. We note that these corpora only include songs where either Taylor Swift or Beyonce were the primary singers (not featured artists or collaborative albums) from their full length albums. Beyonce has a few extended plays and both artists have several singles, however, these songs were all eventually included on studio albums or are mainly remixes which are not useful for our purposes.

For Taylor Swift's corpus, if an album has a re-recorded version, then we only included the Taylor's Version songs and vault tracks for that album. This prevents us from double-counting lyrics and having the most up to date versions of the albums included. Additionally, for Beyonce's corpus, we did not include spanish versions of her songs since the lyrics had the same meaning as the original english versions. For both corpora, no remixes or instrumental versions were included–only original

songs– and we used explicit versions rather than clean versions of songs (wherever applicable) as this was more representative of their respective lyricism.

After creating our corpora, we tokenized the lyrics and removed numbers, punctuations, and symbols so that we only retained words in each corpus. All tokens were converted to lowercase to maintain consistency in our analysis. Laslty, we removed the filler lyrics "like", "ooh", "oh", "yeah", "mm", "mhm", "ah", "oh-oh", "eh", "oo-oo-oo", "whoa", "ah-ah", "ha", "eh", "oh-ah", "uh", "woo", "ya ya", "ya", "yaka", "la-la-la", "na", "na na", from each corpus since these lyrics do not carry any meaning and would skew our analysis if they were not removed.

2.1 Beyonce's Songs Corpus

This corpus consists of 113 songs from 7 albums, with a total of 53,823 tokens. In table 1 we can see the top 5 songs with the most tokens in this corpus.

Top 5 Most Frequent Songs: Beyonce's Corpus

	Tokens	Percent
ring_off.txt	1033	1.919254
mine.txt	934	1.735318
upgrade_u.txt	888	1.649852
video_phone.txt	838	1.556955
baby_boy.txt	795	1.477064
Table 1		

"Ring off", "Mine", and "Upgrade U" are the three largest songs Beyonce has written token-wise, with 1,033, 934, and 888 tokens respectively and accounting for approximately 5.3% of the entire corpus. In table 2, we can examine the proportion each of her albums contribute to the entire corpus.

Proportion of Beyonce's Albums in Beyonce's Corpus

Album	Tokens	Percent
Dangerously In Love	9154	17.013921
Beyonce	8700	16.170102
Renaissance	8685	16.142223
I Am ... Sasha Fierce	7999	14.867201
B'Day	7733	14.372805
4	6732	12.512313
Lemonade	4800	8.921436
Table 2		

Her debut album, "Dangerously in Love", makes up the largest proportion of the corpus, with 9,154 tokens, while Lemonade has the smallest proportion having only 4,800 tokens.

2.2 Taylor Swift's Songs Corpus

This corpus consists of 198 songs from 10 albums, with a total of 74,174 tokens. In table 3 we can see the top 5 songs with the most tokens in this corpus.

Top 5 Most Frequent Songs: Taylor Swift's Corpus

	Tokens	Percent
All_Too_Well_10_Minute_Version.txt	981	1.3225659
I_Wish_You_Would.txt	733	0.9882169
End_Game.txt	692	0.9329415
Out_Of_The_Woods.txt	642	0.8655324
Getaway_Car.txt	565	0.7617224
Table 3		

"All Too Well (10 Minute Version)", "Out of the Woods", and "Endgame" are the three biggest songs in the corpus, with 981, 733, and 692 tokens respectively and accounting for approximately 3.24% of the entire corpus. In table 4, we can examine the proportion each of her albums contribute to the entire corpus.

Proportion of Taylor Swift's Albums in Taylor Swift's Corpus

Album	Tokens	Percent
Red Taylor's Version	20648	24.530140
Speak Now Taylor's Version	9528	11.319410
1989 Taylor's Version	8606	10.224060
Fearless Taylor's Version	8772	10.421270
Midnights	7304	8.677264
Reputation	7186	8.537078
Lover	6531	7.758928
Evermore	6193	7.357379
Folklore	5223	6.205004
Taylor Swift	4183	4.969468
Table 4		

"Red (Taylor's Version)" makes up the largest proportion of the corpus, with 20,648 tokens, while her debut album, "Taylor Swift", has the smallest proportion having only 4,183 tokens.

3 Methods

To address our first research question regarding the difference in parts-of-speech usage between the two discographies' lyrics, we will use a udpipe english model that will tag the tokens in each corpus with what parts-of-speech they belong to. We can then perform frequency and keyness analysis to observe the differences in the prominence of various parts-of-speech tags between the corpora. We will also conduct a principle component analysis for both corpora to identify which parts-of-speech variables have the most influence over each artist's writing.

To address our second research question regarding the context of themes, we can create keywords in context tables to observe the context in which the themes of love and hate are discussed in the songs. Additionally, we can tag the tokens in our corpus using the Biber factor loadings for multidimensional analysis. This will allow us to compare the two discographies based on what factors

are present in each album. Lastly, to address our third research question regarding the change of emotional valence throughout each artist's career, we can conduct a brief sentiment analysis for each corpora and plot the change of emotional valence over time.

4 Results

4.1 Parts-of-Speech

We begin our investigation with a rudimentary analysis of the most frequent words used by each singer. Table 5 displays the top five most frequent tokens found in Taylor Swift's corpus and table 6 shows this for Beyonce's corpus as well.

Most Frequent Tokens in Taylor Swift Corpus

feature	frequency	rank	docfreq
i_prp	4291	1	196
you_prp	3929	2	192
the_dt	2572	3	197
and_cc	2001	4	194
it_prp	1556	5	187
Table 5			

The top five most frequently tokens used by Taylor Swift are "i", "you", "the", "and", and "it" – three of which are personal pronouns and two of which are either a coordinating conjunction or a determiner. The token "i" is occurs the most frequently out of all of the tokens in this corpus, with an absolute frequency 4,291 across 196 songs. This may be indicative of Taylor Swift's personal narrative writing style as she often sings about stories emulating from her personal experiences.

Most Frequent Tokens in Beyonce corpus

feature	frequency	rank	docfreq
i_prp	3232	1	113
you_prp	3019	2	112
me_prp	1371	3	106
the_dt	1368	4	108
it_prp	1228	5	105
Table 6			

Table 6, shows similar results for Beyonce's corpus compared to Taylor Swift's corpus. The top five most frequent tokens in Beyonce's corpus also consist of "i", "you", "the" and "it", with the exception of "me" instead of "and". The token "i" again occurs the most frequently, with an absolute frequency of 3,232 across all 113 of her songs. As was the case with Taylor Swift, Beyonce also writes about her personal experiences in her songs so this may explain the similarity in these results.

Proportion of Parts-of-Speech Tags in Each Corpus

POS	Beyonce	Taylor
Adjective	5.98	4.83
Adverbs	8.78	10.08
Coordinating Conjunctions	2.01	3.64
Determiners	6.54	7.32
Nouns	17.87	15.73
Particles	1.13	0.87
Pronouns	22.54	20.75
Verbs	16.35	17.36

[1] Table 7

As depicted by table 7, further investigation into parts-of-speech frequencies reveals that in Beyonce's corpus adjectives, nouns, particles, and pronouns occur more frequently than in Taylor Swift's corpus. In contrast, adverbs, coordinating conjunctions, determiners, and verbs occur more frequently in Taylor Swift's corpus.

Next, we can perform keyness analysis to identify what tokens in each corpus occur at statistically significant frequencies based on log-likelihood values. For this analysis, our null hypothesis is that there is no difference in frequencies of tokens between the two corpora, i.e. no token has a keyness value greater than or equal to 3.84 for significance level of a p-value less than or equal to 0.05.

Keyness Table: Taylor Swift Corpus as the Reference

Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.4_Tar	Per_10.4_Ref	DP_Tar	DP_Ref
and_cc	262.7772	0.95569	0	2001	758	261.30903	134.72921	0.22014	0.29133
was_vbd	232.8050	2.01467	0	583	106	76.13351	18.84076	0.50997	0.71625
were_vbd	168.2857	3.06510	0	262	23	34.21438	4.08809	0.64077	0.87242
would_md	111.9545	2.24395	0	245	38	31.99436	6.75423	0.66845	0.80873
've_vb	110.4261	4.05774	0	136	6	17.76013	1.06646	0.85340	0.93800

Table 8

Table 8 displays the top five tokens with the highest keyness when using Taylor Swift's corpus as the reference and Beyonce's as the target. We see that these tokens are "and", "was", "were", "would", and "'ve" with keyness values of approximately 262.8, 232.8, 168.29, 112, 110.4 respectively

Keyness Table: Beyonce Corpus as the Reference

| Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.4_Tar	Per_10.4_Ref	DP_Tar	DP_Ref
like_vbp	187.0821	3.09047	0	194	31	34.48215	4.04827	0.77144	0.91125
let_vb	186.0743	2.10297	0	303	96	53.85613	12.53656	0.59729	0.75746
uh_uh	152.3347	2.83499	0	173	33	30.74954	4.30944	0.85897	0.96204
la_nn	122.0435	2.58226	0	154	35	27.37242	4.57062	0.95680	0.98232
that_dt	116.0110	1.10821	0	491	310	87.27182	40.48266	0.51863	0.60231

Table 9

Table 9 displays the top five tokens with the highest keyness when using Beyonce's corpus as the reference and Taylor Swift's as the target. We see that these tokens are "like", "let", "uh", "la", and

"that" with keyness values of approximately 187.1, 166.1, 152.3, 122, and 116 respectively. In both tables we observe that all of the keyness values are greater than the threshold value of 3.84, thus, we can reject the null hypothesis for both situations and can conclude that at least these tokens have differences in use between the two corpora.

Since we rejected the null hypothesis, we can conclude that there do exist differences in the frequencies between the tokens in each discography. We additionally note that Taylor Swift's corpus mainly had verbs, a coordinating conjunction, and a determiner for the top five key tokens. On the other hand, Beyonce's corpus has verbs and a determiner as well, however, it also has a noun and an interjection in contrast to Taylor Swift. These findings could be investigated further to determine what context these differing tokens are used in between the two artists.

Finally, we will conclude this subsection with a brief Principle Component Analysis (PCA) biplot, highlighted in figure 1.

Figure 1: PCA Biplot of Both Discographies
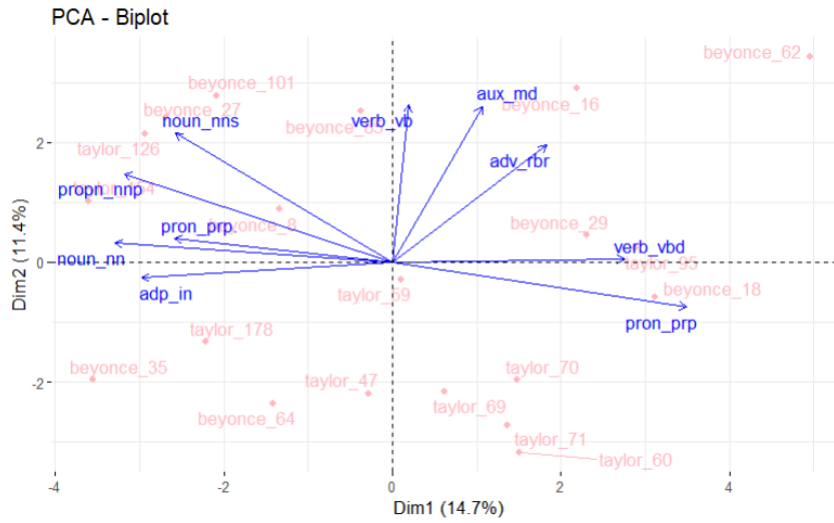
In this biplot, Beyonce's songs are distributed evenly across both dimensions, whereas, none of Taylor Swift's songs have positive scores for both dimension 1 and dimension 2 simultaneously. Neither artist's songs particularly cluster together, so further investigation will be needed to elucidate any potential differences between the principle components of both corpora. Similar to previous observations, we observe that nouns, pronouns, and prepositions, have the strongest influence over songs from both copora. Auxillary modals, adverbs, and past tense verbs also have an influence over Beyonce's songs.

4.2   Context of Themes

In this subsection, we will explore how the specific themes of love and hate are presented by each singer. We can explore this with keywords in context tables for each theme, that highlight the five words that come before and after each theme's keyword for a sample of songs by each artist. We note that although the tables display a sample of songs, the analysis is evaluated based on the entirety of the relevant songs in each corpus.

Keyword in Context of "Love": Taylor Swift

docname	pre	keyword	post
Picture_To_Burn.txt	perfect fantasy i realise you	love	yourself more than you could
Back_To_December.txt	you gave me all your	love	and all i gave you
Mine.txt	wondering why we bother with	love	if it never lasts i
The_Story_Of_Us.txt	d say you d rather	love	than fight so many things
When_Emma_Falls_In_Love.txt	when emma falls in	love	she paces the floor closes
Love_Story.txt	the princess it s a	love	story baby just say yes
All_You_Had_To_Do_Was_Stay.txt	you always want back the	love	they gave away and people
Bad_Blood.txt	it used to be mad	love	so take a look what
Blank_Space.txt	you look my next mistake	love	s a game wanna play
This_Love.txt	this love is good this	love	is bad this love is

Table 10

Keyword in Context of "love": Beyonce

docname	pre	keyword	post
be_with_you.txt	baby baby i	love	the way you make me
crazy_in_love.txt	n't understand just how your	love	can do what no one
dangerously_in_love.txt	i	love	you i love you i
naughty_girl.txt	i	love	to love you baby i
flaws_and_all.txt	do n't know why you	love	me and that s why
why_dont_you_love_me.txt	baby why do n't you	love	me why do n't you
end_of_time.txt	you at night i will	love	you until the end of
i_care.txt	used to be so in	love	used to be so in
love_on_top.txt	you re the one i	love	you re the one i
drunk_in_love.txt	shit happen baby drunk in	love	we be all night last

Table 11

In table 10, we see that in Taylor Swift's songs, "love" seems to appear in context of a crush, heartbreak/breakup, longing/anxiety, and describing relationships. In table 11, we see that in Beyonce's songs "love" is presented in the context of sexual situations, describing relationships, and heartbreak. Beyonce has a more mature discussion of love whereas Taylor Swift sings about budding love that is suitable for a younger audience. Both singers have several songs where they sing about the heartbreak in the aftermath of falling in love.

Keyword in Context of "hate": Taylor Swift

docname	pre	keyword	post
Picture_To_Burn.txt	and by the way i	hate	that stupid old pickup truck
Castles_Crumbling.txt	me never wanted you to	hate	me my castle s crumbling
Dont_You.txt	really wish that i could	hate	you i ve tried but
We_Are_Never_Ever_Getting_Back_Together.txt	a day i say i	hate	you we break up you
Bad_Blood.txt	n't hate you but i	hate	to critique overrate you these
Shake_It_Off.txt	play and the haters gonna	hate	hate hate hate hate haters
Gorgeous.txt	so cool it makes me	hate	you so much i hate
I_Forgot_That_You_Existed.txt	i did it is n't	hate	it s just indifference it
You_Need_To_Calm_Down.txt	about all the people you	hate	cause shade never made anybody
The_Archer.txt	spite my face then i	hate	my reflection for years and

Table 12

Keyword in Context of "hate": Beyonce

| docname	pre	keyword	post
i_cant_take_no_more.txt	in your mouth what you	hate	me for i m so
ring_the_alarm.txt	touching her is what i	hate	most i do n't want
broken_hearted_girl.txt	there are times when i	hate	you cause i ca n't
dance_for_you.txt	show you how much i	hate	being apart yes show you
jealous.txt	no no no and i	hate	you for your lies and
cozy.txt	juicy you hate me you	hate	me cause you want me

Table 13

In table 12, we see that in Taylor Swift's songs, "hate" comes up when she sings about an ex-boyfriend, a breakup, revenge, and her anti-fans. In contrast, in table 13 we can observe that Beyonce sings about "hate" in regards to an ex-boyfriend when discussing a failed relationship and jealousy. We note that Taylor Swift has several more songs with the theme of "hate" than Beyonce does. Both singers sing about hating their ex-partners from relationships that were traumatic for them, however, Taylor Swift also discusses her enemies and anti-fans in addition to this.

4.3   Dimensional Analysis

In this section we will use the Biber loadings to conduct dimensional analysis to identify the different factors influencing each discography.

8

Figure 2: (+) Descriptive v.s. (-) Informative
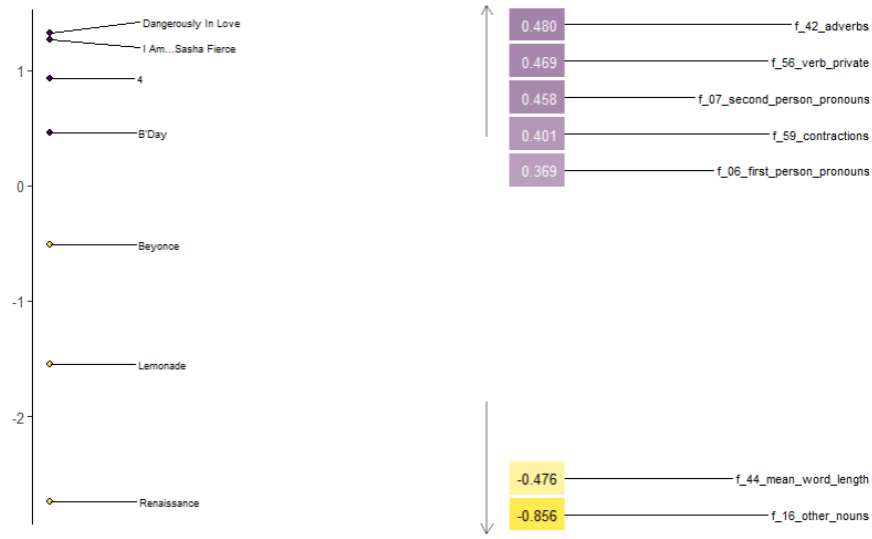
In figure 2, we can observe that Beyonce's albums are equally spread across this dimension, with the exception of one cluster of albums at the positive factor loadings side. These four albums have high scores for second and first pronouns, adverbs, private verbs, and contractions, indicating that they are more descriptive than the other albums. The remaining three albums are distributed evenly along the negative factor loadings, indicating they are more informational than the four albums that are clustered together.

Figure 3: (+) Involved, Narrative v.s. (-) Informative, Descriptive
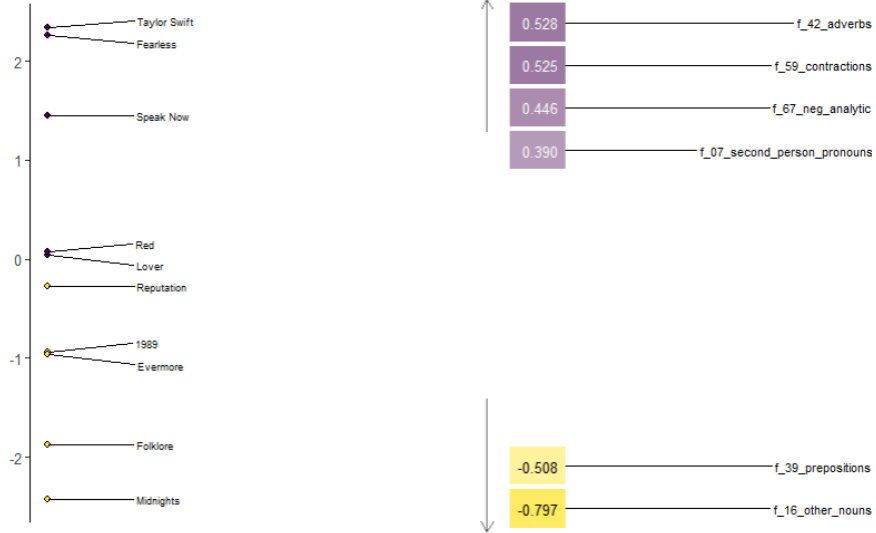
In figure 3, we can observe that Taylor Swift's albums have are varied across this dimension. There is one cluster of albums with positive factor loadings for this dimension. indicating that they are more narrative than the other albums. Two albums are clustered on the negative end of this dimension,

indicating they are more informational and have more descriptive lyrics than the albums on the positive end. The third cluster of albums in the middle of the dimension with scores ranging from -1 to 0 have an even mix of characteristics from both ends of this dimension.

Overall, Taylor Swift's songs range from informative and descriptive to narrative in their storytelling nature, whereas, Beyonce's songs range from descriptive to informative respectively. We also note that for the positive factor loadings, three of the factors are the same as Beyonce's positive factore loadings (adverbs, contractions, and second person pronouns) and one of the negative factor loadings are the same as well for both singers (other nouns).

4.4   Sentiment Analysis

We will conclude our analysis for this report with a brief sentiment analysis plot displayed in figure 4 to compare emotional valence over time for the two singers' lyrics.
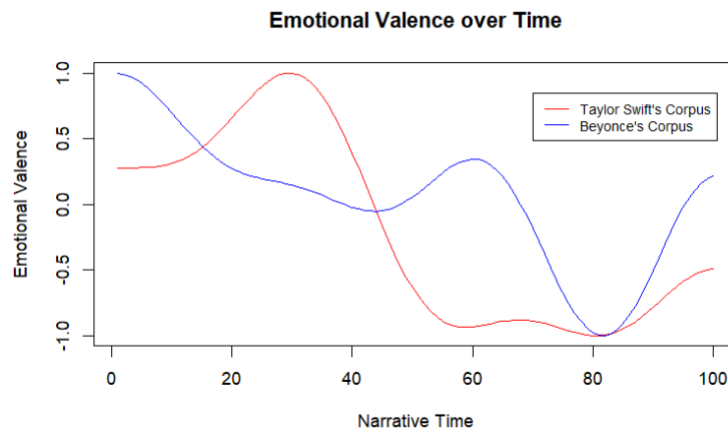

Figure 4

In figure 4, we can observe that Taylor Swift's lyrics begin somewhat positive, become increasingly positive, then drastically become more negative. Her newer songs have more negative emotional valence scores than her older songs, however, her emotional valence scores have been improving/increasing in recent narrative time. Similarly, Beyonce's lyrics start off with incredibly positive scores, which continuously decrease, then slightly increase before decreasing again, and end with an increase in emotional valence scores. Although both artists seem to have decreasing emotional valence scores throughout their discography, Beyonce maintains positive scores for a majority of her songs, whereas, a majority of Taylor Swift's songs received negative scores.

5   Discussion

5.1   Conclusions

To recap our investigation, first we observed the differences in the frequencies of the tokens and the parts-of-speech tags between the two corpora, performed keyness analysis and principle component analysis, observed key words in context for "love" and "hate", conducted multidimensional analysis and ended with sentiment analysis.

We found that both Beyonce and Taylor swift had a majority of the same top five most frequent tokens in each corpus. Additionally, Beyonce used more adjectives, nouns, particles, and pronouns, whereas, Taylor Swift used more adverbs, coordinating conjunctions and determiners. We

also observed that there was a statistically significant difference in the frequencies of tokens when using both corpora as references and targets interchangebly for keyness analysis. Moreover, through a PCA biplot, we observed that similar parts-of-speech tags (nouns, pronouns, and propositions) influence both artists's writings, however, Beyonce's writing has additional influence from auxillary modals, adverbs, and past tense verbs in comparison to Taylor Swift. These findings provide a foundation for a future project for a deeper investigation into what differentiates the lyrics of these two artists in regards to frequency and parts of speech.

We also found that both Beyonce and Taylor Swift address the theme of "love" in context of heartbreaks, breakups, and when describing relationships. However, Taylor Swift often sings about "love" in context of crushes, anxiety, whereas, Beyonce sings about "love" in relation to sexual acts/desires. This makes Taylor Swift's songs more targeted for a younger audience in contrast to a mature audience targeted by Beyonce's songs. Both artists also sing about "hate" in regards to past failed relationships and ex-boyfriends. We note that Taylor Swift sings about hate more often than Beyonce does and she also sings about it in context of enemies, revenge, and anti-fans. This makes some of her songs a bit more negative than what Beyonce presents. Through multidimensional analysis we found that both artists have similar positive factor loadings and negative factor loadings. Subtle differences in these factor loadings highlight that Taylor Swift's albums are either descriptive or informative in their story-telling while Beyonce's albums are either narrative or descriptive with respect to positive and negative dimension scores.

Sentiment analysis shows that both artists had positive emotional valence in their songs around the time of their debut albums, however, this emotional valence decreased in score for both discographies as they released more songs, indicating an increase in negative. We note that in recent times, both artists exhibit an increase in emotional valence scores, on their way to becoming more positive in their writing. Overall, Taylor Swift has more negativity in her discography than Beyonce does.

We conclude that their are a few similarities between the two artists' writings, however, as mentioned earlier in this subsection, there are a few differences that are worthy of future investigation for a larger and more substantive project.

5.2   Limitations

We note that the lack of including collaborations, original soundtrack songs/theme songs for movies and TV shows, and features in each corpus limits the depth of our analysis. It is plausible that these excluded songs could highlight additional differences between the artists that we did not observe in this analysis, however, for the consistency of our corpora, we did not include these songs.

Additionally since this project is intended to be a "proof of concept" for a much larger project, we did not perform topic modelling. Instead, we investigated the usage of two specific themes exhibited in a majority of pop songs that we essentially cherry picked to analyze. This limited our analysis to examining keywords of each theme in context rather than a more in-depth exploration of specific themes present in each discography. The keywords in context tables are also not as substantive without stemming since we were not able to map similar words together (e.g. "lovers" and "lover" have the same stem but were not grouped together in our analysis).

Lastly, we note the limitation in our key word analysis due to the scarcity of context in the table. We only observed the 5 words appearing before and after "love" and "hate" in relevant songs from each corpus to determine how they were being used by each artist, however, we may need more complete sentences for better accuracy of our discernment and beliefs about the context in which these themes were being used in.

5.3 Future Outlooks

Future analysis could employ stemming and topic modelling. Stemming will allow us to group words with the same stem together for better analysis of the context in which words associated with certain themes are used in each artist's songs. Topic modelling will allow us to determine which themes are present in each discography to centralize and focus our comparisons towards these specific themes rather than themes we cherry picked. Once we determine these themes, we can investigate the usage of words related to these themes in each corpus using collocations and keywords in context methods similar to as we did with the cherry picked themes. Future analysis could also explore the usage of parts-of-speech by creating PCA plots for each artist individually, by examining unique noun phrases in each corpus, and creating more keyness tables for various parts-of-speech tags to substantiate our investigation into the differing usage of various parts-of-speech between the corpora. It may also be possible to perform regression analysis to determine how parts-of-speech relate to n-grams in each artist's lyrics.

6 References

"Alphabetical List of Part-of-Speech Tags Used in the Penn Treebank Project." Penn Treebank P.O.S. Tags, Pen Treebank Project, 2003, `www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html`.
"Beyonce." Genius, ML Genius Holdings , `genius.com/artists/Beyonce`. Accessed 19 Nov. 2023.
"Taylor Swift." Genius, ML Genius Holdings , `genius.com/artists/Taylor-swift`. Accessed 19 Nov. 2023.