36-668: Coffee Break Experiment 1

Arsh Gupta

Contents

1 Introduction

Adele is a widely revered global artist who has been the recipient of multiple awards. Her songs are characterized by a combination of soulful, deep, and heartfelt emotions, the themes of which are shared across her four albums: 19, 21, 25, and 30. Shared among these is the theme of **love**, which has evolved in the way Adele characterizes it across her songs. The purpose of this report is to analyze the was that Adele's conceptualization of the theme of love has evolved across her different albums and what insights that gives us about the manner in which she communicates about this topic.

2 Data

The data set used for this report includes a corpus consisting of 46 text files. Each of those text files includes lyrics to each of Adele's 46 songs spanned across her four albums. The data was manually collected from the web through Google and compiled into a corpus using the `quanteda` library in R.

3 Methods

3.1 Data Pre-Processing

There were multiple steps of data processing that were adopted before conducting the analysis which have been described in the next section. We have also obtained summary statistics of our data at multiple steps for different albums to better understand the features of each album. Tables 1, 2, 3, and 4 show a partial summary of the four albums listing the number of tokens and overall sentences in each song of that particular album.

The reason we perform data pre-processing is to prepare our text in a manner that is usable for subsequent analysis.

Table 1: Partial summary of Album 1: 19

Text	Types	Tokens	Sentences	type	album	song_rank
song_1_1.txt	107	197	1	song	1	1
song_1_10.txt	110	370	1	song	1	10
song_1_11.txt	108	331	17	song	1	11
song_1_12.txt	100	246	3	song	1	12
song_1_2.txt	160	439	4	song	1	2
song_1_3.txt	89	325	27	song	1	3

Table 2: Partial summary of Album 2: 21

Text	Types	Tokens	Sentences	type	album	song_rank
song_2_1.txt	147	625	1	song	2	1
song_2_10.txt	33	185	1	song	2	10
song_2_11.txt	130	384	3	song	2	11
song_2_2.txt	113	426	1	song	2	2
song_2_3.txt	91	259	1	song	2	3
song_2_4.txt	95	219	10	song	2	4

Table 3: Partial summary of Album 3: 25

Text	Types	Tokens	Sentences	type	album	song_rank
song_3_1.txt	134	429	4	song	3	1
song_3_10.txt	119	297	5	song	3	10
song_3_11.txt	116	276	1	song	3	11
song_3_2.txt	96	442	3	song	3	2
song_3_3.txt	114	440	1	song	3	3
song_3_4.txt	133	405	4	song	3	4

Table 4: Partial summary of Album 4: 30

Text	Types	Tokens	Sentences	type	album	song_rank
song_4_1.txt	87	128	2	song	4	1
song_4_10.txt	130	413	4	song	4	10

Text	Types	Tokens	Sentences	type	album	song_rank
song_4_11.txt	130	354	2	song	4	11
song_4_12.txt	141	396	3	song	4	12
song_4_2.txt	97	213	1	song	4	2
song_4_3.txt	190	489	9	song	4	3

The data is then further processed to remove punctuation, symbols, and numbers from the tokens obtained. All the tokens are also changed into lowercase to ensure consistency and avoid redundancy. As a next step, we have combined all multi-word expressions into one token using the mwe list called `multiword_expressions` as found in the `cmu.textstat` library.

3.2 Summary Statistics

At this point, we have combined all tokens from the first two albums and last two albums into a single object leaving us with a total of two token objects since we will be analyzing the conceptualization of love as found in Adele's first two albums versus the last two albums.

3.2.1 Token Frequency

At this point, we compare the frequency of most common tokens across different albums. In the Results section, Table 5 shows the most common tokens across all albums, and Tables 6 and 7 show the most frequent tokens specifically within the first two and last two albums respectively.

We compute the token frequency because it is the most rudimentary way to assessing the common themes and subjects across Adele's songs.

3.2.2 Dispersion Measures

As a next step, we compute the dispersion measures using AF and DP. In the Results section, Table 8 shows the dispersion measures of most common tokens across all albums, and Tables 9 and 10 show the dispersion measures for most frequent tokens specifically within the first two and last two albums respectively.

We compute the dispersion measures because it enables us to see if certain tokens that appear more commonly are perhaps localized or evenly distributed. This will pave the way for further research questions as to whether certain songs focus on certain themes or not.

3.3 Collocational Network Plot

As the final step in our analysis, we compute collocates for each token in the two corpora containing songs from first two and last two albums respectively. These collocates have been computed using the mutual information statistic, and once we obtain that, we plot two graphs; one for token **love** and one for token **heart** to understand how the information conveyed by Adele about those two words has changed across the two sets of albums and whether there are any commonalities.

4 Results

4.1 Tokens Frequency

Tokens **i**, **you**, and **the** are the most common across corpora containing first two and last two albums. Token **i** occurs much more frequently in the last two albums versus the first two, suggesting that there is a greater

shift towards personification in the songs from Adele's last two albums.

Table 5: Most Frequent Tokens Across All Albums

Token	Frequency
i	749
you	569
the	427
me	370
to	360
it	305

Table 6: Most Frequent Tokens Across First Two Albums

Token	Frequency
i	340
you	319
the	186
it	177
and	158
to	157

Table 7: Most Frequent Tokens Across Last Two Albums

Token	Frequency
i	409
you	250
the	241
me	213
to	203
my	129

4.2 Dispersion Measures

None of the values of dispersion measures are surprisingly high, suggesting that the most common tokens across the two corpora are even distributed across Adele's songs.

Table 8: Dispersion Values of Most Frequent Tokens Across All Albums

Token	AF	DP
i	749	0.22
you	569	0.27
the	427	0.26
me	370	0.35
to	360	0.24
it	305	0.42

Table 9: Dispersion Values of Most Frequent Tokens Across First Two Albums

Token	AF	DP
i	340	0.23
you	319	0.25
the	186	0.29
it	177	0.46
and	158	0.22
to	157	0.27

Table 10: Dispersion Values of Most Frequent Tokens Across Last Two Albums

Token	AF	DP
i	409	0.21
you	250	0.29
the	241	0.24
me	213	0.32
to	203	0.22
my	129	0.28

4.3 Collocational Network Plot

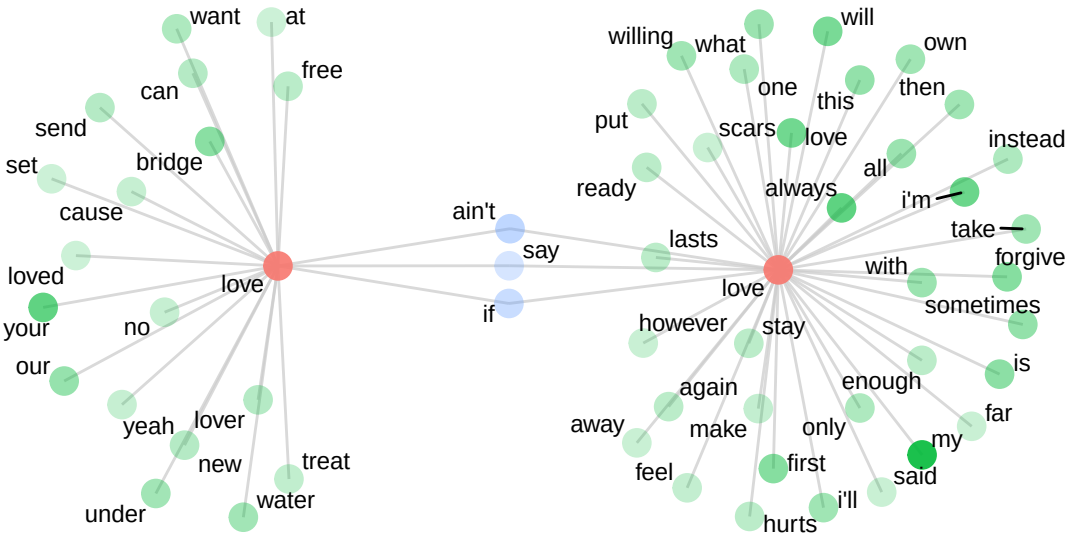
We obtain two collocational network plots. The left and right clusters of tokens in Graph 1 show the collocates having highest MI values for token **love** in songs from the first two and last two albums respectively. Graph 2 shows a similar relationship but for the token **heart**. We choose these two tokens because seem like reasonable choice of words that Adele would use to communicate about love and related themes.

One interesting insight that we observe is that most tokens containing high mutual information about token **love** from the last two albums (scars, forgive, hurts) tend to have somewhat of a negative/painful connotation relative to tokens from the first two albums.

On the other hand, collocates for token **heart** across all albums seem to be have more sad themes (cried, wish, blood, breaking, cry, heal, stain) suggesting that the occurrence of **heart** throughout all of Adele's songs happens within a sad or painful context.

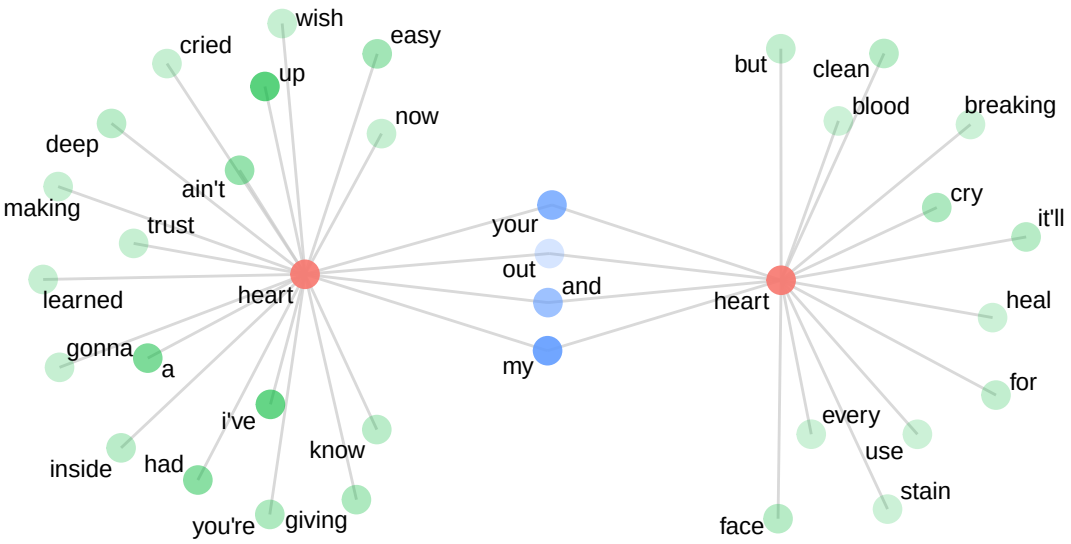
We see that for the token **love**, songs from first two and last two albums share three common collocates, namely: **ain't**, **say**, and **if**.

Graph 1: Collocates for token love across first two and last two albums



Similarly, for the token **heart**, songs from first two and last two albums share four common collocates, namely: **your**, **out**, **and**, and **my**.

Graph 2: Collocates for token heart across first two and last two albums



5 Discussion

5.1 Conclusions from findings

Our analysis does yield us some interesting insights about how some shared themes in Adele's songs have evolved across the different albums.

The fact that most collocates for **love** from the last two albums convey a sad context and that collocates for **heart** from almost every album convey that context suggests that Adele's conceptualization of love has progressed from being less to more sad across various songs in her four albums.

From the first collocational network plot, we see that in songs from her first two albums, the highest mutual information about token **love** is contained in **ain't**, **say**, and **if**.

- The first one, **ain't** suggests that the theme is conveyed through some sort of a comparison or contrast, where Adele establishes a parallel between love and something that shouldn't be.
- The second collocate, **say**, suggests something actionable, with an emphasis on vocalizing the presence of love.
- The third one, **if** hints at some conditional implication that would be contingent upon the presence or existence of love.

From the second collocational network plot, we see that in songs from her first two albums, the highest mutual information about token **heart** is contained in **your**, **out**, **and**, and **my**.

- The collocates, **your** and **my** suggests that the occurrence of token **heart** is within a dependency context, possibly with another person.
- The second one, **out** might be referring to a phrase such as "put your heart out" or something similar, suggesting that there is an underlying theme of encouraging expression, emotion, and vulnerability.

5.2 Scope for further study and limitations

While this analysis is a good first step in analyzing the underlying themes within Adele's songs, there is definitely a scope of further study. The level of analysis conducted here is pretty surface level and does not account for external factors that might influence the manner in which certain themes are conveyed in Adele's songs. Some variables that might influence this are the socio-economic climate of the time her album was released, any possible intended purpose of the album, etc.