International Journal of Modern Languages and Applied Linguistics
e-ISSN: 2600-7266

A Corpus-driven Analysis of Taylor Swift's Song Lyrics

Faruq Ahmad Kendong[1], Afifah Sakinah Daud[2], Siti Aeisha Joharry[3]*

[1, 2 & 3] Akademi Pengajian Bahasa, Universiti Teknologi MARA

[1] faruq.ahmadk94@gmail.com
[2] 2022276694@student.uitm.edu.my
[3] aeisha@uitm.edu.my

**Corresponding author*

Article history:
Received: 15 April 2023 *Accepted: 19 May 2023* *Published: 1 June 2023*

Abstract

Taylor Swift is widely recognized for her song writing skills, known for crafting relatable lyrics, catchy tunes, and personal storytelling that draws from her experiences and emotions. This exploratory study aims to explore the language used in Taylor Swift's song lyrics by analysing the language features, its collocations and the most used lexical bundles. A specialised corpus, the Taylor Swift Corpus (TSC), was compiled with a total of 189 song lyrics from 10 Taylor Swift albums, from her debut album - Taylor Swift (2004) to the latest release - Midnights (2022). TSC consists of song lyrics of different genres, including country, country rock, pop, country pop, folk and folk pop. #LancsBox 6.0 software was utilised for generating the statistics of the corpus, where the results were analysed further in order to fulfil the proposed research objectives in this research. Various corpus analyses were carried out including frequency analysis, collocational patterns analysis and examining the concordance lines. Findings revealed Taylor Swift's writing styles that are personalised in nature through use of the personal pronouns 'I' and 'you', followed by salient references to 'time' and her reflections on events via the cognitive verb 'know'. Future studies could investigate other properties of the artist's writing such as metaphorical or figurative use of language.

Keywords: *corpus linguistics, lexical bundles, song lyrics, Taylor Swift, Taylor Swift Corpus (TSC)*

1.0 Introduction

While corpus research on music or song lyrics is nothing new, similar approaches to a single artist's discography is very unlikely. This is mainly due to restrictions that exist for corpus research, where a large collection of texts usually produces better findings (Ghane & Farahani, 2021) and that corpus research for song lyrics is usually compiled from various singers using music charts as inclusion criteria (Eiter, 2017; Nishina, 2017; Muhammad et. al, 2022).

Many studies that investigated chart-topping songs in various areas include metaphors (e.g., Climent & Coll-Floritt, 2019), relationship between linguistic items and popularity of songs (Nishina, 2017), and profanity in song lyrics (Muhammad et. al, 2022) to name a few. Chart topping songs often feature multiple contributing writers that assist singers to expedite the creation of songs to match the fast-paced music industry. As Sutherland (2017) argues, only 5 songs in the Top 100 were written by a single songwriter while 13 percent of the biggest 100 hit songs credit 8 or more songwriters. This means that a single songwriter's writing style and linguistic ability would not be representative from the large involvement of contributing writ-

ers. Therefore, a study that is specific to examining one songwriter's body of work requires a specialised corpus to be created, presupposing that the author/songwriter should be one with a sufficiently expansive body of work that is not dependent on multiple contributing writers to achieve an appropriate level of representativeness.

1.1 Taylor Swift's song lyrics as a specialised corpus

When examining songwriters in the current music landscape, few fit the criteria, lacking sufficient discographies for an accurate corpus while others may feature multiple contributing writers that would not be representative of the target songwriter. Taylor Swift's discography (i.e., the history of her recorded music) on the other hand, features songs that were mostly written by her. Furthermore, the greatest number of contributors involved with her songs up until her fifth album, '1989', were four people for the songs, Untouchable in her second album, 'Fearless' and Style in '1989'. Her songs would usually feature one or two contributing writers, some examples are shown in Table 1.1.

Table 1.1: Co-writers in Different Albums

Track	Co-writer(s)	Album
Teardrops on My Guitar	Liz Rose	Taylor Swift
Tim McGraw		
You Belong with Me	Liz Rose	Fearless
Breathe	Colbie Caillat	
Everything Has Changed	Ed Sheeran	Red
I Knew You Were Trouble	Max Martin, Shellback	
Blank Space	Max Martin, Shellback	1989
Clean	Imogen Heap	

Her subsequent albums feature songs with four or more contributing writers; the songs and their respective albums are listed in Table 1.2.

Table 1.2: Songs with More Than 4 Contributing Writer

Track	Co-writers(s)	Album
End Game	Martin, Shellback, Sheeran, Nayvadius Wilburn	Reputation
Look What You Made Me Do	Antonoff, Richard Fairbrass,	

	Fred Fairbrass, Manzoli	
Lavender Haze	Antonoff, Zoe Kravitz, Mark Spears, Jahaan Sweet, Sam Dew	Midnights
Karma	Antonoff, Spears, Sweet, Keanu Torres	

She is also an artist with a body of work that spans nearly two decades with accolades that justify her significance in the music industry since her debut in 2004. She has won a numerous number of awards, amongst them are the Top Female Artist in 2022 (Billboard, 2022), 40 American Music Awards, the most won by any artist (AMAs, 2008-2022), the Most Streamed Music Artist of 2022 (NBC, 2022), the Longest Reigning Act of Billboard Artist 100 (Billboard, 2022) and the Nashville Songwriters Association International's Songwriter-Artist of the Decade Award (NME, 2022).

According to Kreyer and Mukherjee (2007), the genre of pop lyrics has not been studied in-depth and that pop song lyrics are not found in any standard reference corpora including the British National Corpus (BNC). It can be argued therefore that corpus linguistic studies in the past rarely examined pop songs or their lyrics and in turn, the present study presents a corpus-based study on pop song lyrics, specifically Taylor Swift's song lyrics that are mostly from the pop genre, followed by country, rock and folk music.

Also, at the time of writing, there was no available corpus of Swift's song lyrics from her first album – 'Taylor Swift' (2004) to her latest release – 'Midnights' (2022), and the only Taylor Swift corpus found was compiled from Swift's first self-titled album until her 'Reputation' (2017) album (Boonjoon, 2018). This might be because researchers rarely use this approach in analysing song lyrics, and usually they will opt for other approaches and methods such as critical discourse analysis (Alek et. al, 2020), non-participant observation (Setiawati & Maryani, 2018; Syafar & Asty, 2022), and auditory analysis and interviews (Lyon, 2019). Hence, this research gap has led us to compile the lyrics from all Swift's albums that would be more representative of her song writing career.

1.2 Research objectives

The aim of this study is to investigate Taylor Swift's song lyrics using corpus linguistic analysis. Thus, the following research objectives are formed:

1. To identify the language patterns and linguistic items in Taylor Swift's writing
2. To examine the most overused lexical bundles (or repeated strings of words) in her writing

2.0 Literature Review

2.1 Corpus linguistics

One of the ways to study language is through examining language used in naturally-occurring written or spoken texts (regarded as "corpora"; corpus for singular). Termed as 'Corpus linguistics', McEnery and Hardie (2012) define this as "the study of language data on a large scale - the computer-aided analysis of very extensive collections of transcribed utterances or written texts". In their book, they highlight that the field of study focuses upon a set of procedures, or methods, for studying language (McEnery & Hardie, 2012: p.1) and that it is debatable whether the characterisation of corpus linguistics is considered a theory in its own right or as a methodology. In this paper, corpus linguistics is defined as a methodology, following McEnery and Hardie as well as linguists who share similar views (e.g., Gries, 2020; Tognini-Bonelli, 2001; Mahlberg, 2010). More importantly, corpus linguistics can be considered as Lee (2008: p.87) terms it: "a methodology innovation" because one usually internalises results of corpus data with a set of theoretical positions and beliefs, he or she may have about the language being studied.

There are numerous benefits of using corpus to study language including its ability to provide comprehensive knowledge of linguistic structures that can be facilitated by statistical software that generates different types of information such as word frequencies that are used to either come up with or confirm a concept, theory and rule (Atar & Erdem, 2019). Moreover, Atar and Erdem (2019) added that corpus linguistics is an excellent tool for gathering information about language studies as it allows researchers to analyse an enormous size of data as these data will be stored, compiled and analysed digitally with the use of special software. In addition, Wulff (2015) believed that corpora are not limited to the written form, where in some cases, a certain corpus claims to be a representative of a specific variety of discourse that include spoken data.

Given its methodological nature, corpus approach has been incorporated in a range of linguistic research such as in language teaching, language acquisition, translation studies, discourse analysis, stylistics, metaphor, functional linguistics, World Englishes and many more. While there are various trends and themes in the development of corpus research (see Baker and contributors, 2009 for a collection of contemporary corpus linguistic studies), typical techniques are often found with regard to frequency analysis, co-occurrence of linguistic items as well as close reading of the text to examine language use in context. These are normally referred to as 'keywords analysis' for identifying salient key words or terms in one specified corpus relative to the words/terms occurrence in another (reference) corpus; 'collocations analysis' that presents statistically significant co-occurrences between words in order to examine relationships between words or phrases; and finally 'concordancing' that displays words in context by analysing search words/phrases ordered neatly in the middle with co-text on either side of it/them.

2.2 Language studies on song lyrics

Traditionally, "[s]ong lyrics often begin as lyric poems" (Craven, 2021). A simple search on Google Scholar would list a range of studies on song lyrics published last year, mostly related to language such as investigating metaphors in song lyrics (e.g., Simanjuntak et al., 2022), deixis (e.g., Damayanti, 2023), and figurative language (e.g., Yusnitasari et al., 2022). While these studies concluded that song lyrics are mainly expressive and display personal feelings of the singer-songwriter through use of language, their approach to analysing song lyrics are purely qualitative. In other words, similar research like the ones listed above begin to choose a language feature and examine their manifestations in one – usually popular – song lyric.

Other types of language studies on song lyrics range from those that explore how a semantic analysis on lyrics mined from the Web enabled for a systematic approach to indexing music content (Logan et al., 2004) to pedagogical applications of song lyrics (Hadian, 2015). Although Hadian (2015) found that students responded positively to learning with music (focus on listening), Racette and Peretz (2007) recalled that no significant impact was discovered for using lyrics and songs in vocabulary learning. Another group of language studies on song lyrics employs computational methods, i.e., automatic retrieval or generation of song lyrics can be found in the literature specific to another type of semantic analysis (e.g., Oliveira, 2015; Knees et al., 2005).

2.3 Song lyrics and corpus studies

According to Choi and Downie (2019), the complexity of audio music has drawn many researchers to employ computational methods to better study and understand its different auditory facets while the contents of the lyrics are often excluded from the research. Similarly, the use of corpora to study song lyrics is not a foreign proposition. As Kreyer and Mukherjee (2007) have argued, corpus linguistic studies have generally excluded pop song lyrics from standard reference corpora like the British National Corpus, despite corpora being demonstrated to be a potent research resource for the descriptive features of lyrics. For example, Kreyer (2015) found that pop song lyrics often emphasised and reinforced traditional gender roles as well as stereotypes, with men portrayed as dominating and active in contrast to women as passive. In a similar vein, Krasse (2019) highlighted the difference in linguistic patterns by female artists from the Billboard Hot 100 chart which indicated that women presented themselves independent and strong, in contrast to traditional gender roles and stereotypes.

While these studies have utilised corpora to examine song lyrics, they make use of song charts as selection criteria for the data's inclusion in the corpus. In contrast, corpus-related research on an individual songwriter's collection of work is more limited; exemptions include Yeh et al. (2019) that analysed the boy band, One Direction's albums using corpus-assisted discourse analysis and investigated how love and romance are represented. However, criticisms against examining boy bands' albums are reported in that the

success of a band of singers such as One Direction is highly dependent on the contributions of various songwriters making them (the contributions) crucial in the song writing process (Wolk, 2012), and as a result, selection of albums by any singer or group of singers would need to be screened carefully to ensure that songs selected for analysis are direct representations of the singer-songwriter. Meanwhile, Boonjoon (2018) investigates the linguistic characteristics and genre shift in Taylor Swift's song lyrics, comparing the two periods in her career, the first period considered as country-pop (2006-2013) to her pop period (2014-2017). Boonjoon's findings revealed that there was a shift towards the artist's greater use of colloquial and informal language, pronouns, personal experiences and descriptions of emotions during her pop period compared to the earlier one. These studies signify the capabilities of corpus methods in analysing song lyrics.

3.0 Corpus Description and Methods

According To reiterate, the present research explores Taylor Swift's song lyrics from 2004 to 2022 via methods in corpus linguistics. More specifically, this study employed a corpus-driven approach where word lists are firstly extracted using a corpus tool/software (quantitative) and are used as a starting point followed by close readings of concordance lines, which are qualitative in nature. Before the data can be analysed, a specialised corpus – Taylor Swift Corpus (hereafter TSC), was compiled. As a reminder, this corpus consists of the artist's songs from her self-titled album (Taylor Swift, 2004) to her latest album (Midnights, 2022). A total of 189 song lyrics files were included in the corpus. Table 3.1 shows the portions of the song lyrics from 10 different albums released as well as the total number of words.

Table 3.1: Taylor Swift Corpus (TSC) description

No.	Album Name	Year	Number of Songs	Number of Words
1	Taylor Swift	2004	15	3,016
2	Fearless	2008	25 (*TV: 7)	6,239
3	Speak Now	2010	17	5,065
4	Red	2012	28 (*TV: 10)	7,705
5	1989	2014	16	3,830
6	Reputation	2017	15	4,317
7	Lover	2019	18	4,545
8	Folklore	2020	17	4,248
9	Evermore		17	4,786
10	Midnights	2022	21	5,097
Total			189	48,848

**TV: Taylor's Version*

Since we included all her songs (irrespective of genre) in TSC, it can be argued that the corpus is both representative and balanced of Taylor Swift's song writings. Additionally, we restricted songs chosen to those featured in any of Taylor Swift's albums excluding singles (individually released songs) made and written for movies. This is because songs made for movies often feature external input from the movie directors, making the songs unrepresentative of the singer's lyrical and writing choices.

When obtaining and cleaning text for this corpus, additional considerations were made to further improve accuracy of data, especially repetitions in songs. Although studies like Nishina (2007) have found repetitions like "la la la", "nanana" and "oh oh oh" typical of song lyrics, these phrases function more as fillers than specific meanings to the song. As a result, such features of repetition in the corpus that are used as fillers were removed. However, the chorus of songs which are repeated with the same lyrics were retained up to only two times (and subsequent repetitions of the exact lyrics, i.e., chorus are removed).

Following these considerations, the data was then uploaded to #LancsBox 6.0, which generated 48,848 tokens with 3972 types from the 189 songs. Using the software, a keywords analysis using simple maths was carried out. After that, the collocations (or habitual juxtaposition of a particular word with other word/words with a frequency greater than chance) and concordances of the corpus were identified. Measures that were taken to analyse collocations include restricting the criteria of identification (of collocates) to a t-score of five and significant collocates should occur at least seven times in the corpus. The N-grams feature was also carried out to examine frequent strings of words, also known as multi-word sequences (following Biber, 2009) or lexical bundles (the most frequently recurring sequences of words in a register such as *I don't know if* or *I just wanted to*). As Biber asserts, characteristics of generating these examples of formulaic language are outlined below:

1. It would be based on the actual word forms that occur in the corpus and not lemmas
2. It would be based on analysis of sequences of word forms, with no consideration given to the grammatical/syntactic status of those words
3. It would focus on frequent, recurrent combinations of word forms

(2009: p.281)

Lexical bundles are identified using a frequency-driven approach (Biber, 2009: p.282) and for the purpose of this study, only trigrams (or 3-grams) were investigated. Findings from these analyses are discussed in the following section.

4.0 Findings

4.1 Word frequency analysis

Table 4.1 shows the results of the word frequency analysis. The top 10 most occurring words were mostly pronouns (*you, I, me, and my*) and other functional words like *the, and, to, a, in,* and *it,* where they occurred in more than 82% of the files in the TSC. Although closed class (or functional) words may indicate stylistic in writing, they are considered highly overused (Boyd et al., 2020) and other studies like Hills and Adelman highlighted that many content words (or open class words), such as nouns, lexical verbs, adjectives, and adverbs were found to "have relatively higher concreteness scores than closed class words" (2015: p.48) in popular song lyrics and that they are indicators of trends in song lyrics. In turn, Table 4.2 presents a list of top 10 open class words and their frequencies (in terms of nouns and lexical verbs) in Taylor Swift's song lyrics to be analysed.

Table 4.1 Top 10 Word Frequency List

	Type	**Frequency**	**Range Percent (%)**
1	you	2232	96.83
2	i	2134	98.94
3	the	1810	100.00
4	and	1464	98.41
5	to	902	97.35
6	me	859	92.59
7	a	838	94.18
8	in	759	94.709
9	it	722	86.24
10	my	679	87.30

Table 4.2: Top 10 Open Class Word Frequency List

	Word (Nouns)	**Freq. / Dis. (%)**
1	time	250 / 56.61
2	love	119 / 34.40
3	thing	115 / 37.04
4	night	115 / 36.51
5	baby	99 / 24.34
6	eye	90 / 31.75

	Word (Lexical Verbs)	**Freq. / Dis. (%)**
1	know	475 / 74.07
2	say	334 / 67.20
3	do	285 / 56.08
4	get	280 / 62.96
5	have	266 / 58.73
6	go	246 / 55.56

	Word (Nouns)	Freq. / Dis. (%)
7	way	89 / 29.63
8	day	81 / 26.98
9	girl	72 / 20.63
10	everything	71 / 20.63

	Word (Lexical Verbs)	Freq. / Dis. (%)
7	see	210 / 52.91
8	think	208 / 47.09
9	make	101 / 47.09
10	come	159 / 44.97

Based on Table 4.2, *time* is ranked as the highest noun occurring 250 times across 57% songs, followed by *love* (119; 34%), *thing* (115; 37%), *night* (115; 36.5%), *baby* (99; 24%), *eye* (90; 24%), *way* (89; 29.6%), *day* (81; 27%), *girl* (72; 21%), and *everything* (71; 20.6%). In terms of lexical verbs, *know* was ranked highest with 475 times across 74.1%, followed by *say* (334; 67%), *do* (285; 56%), get (280; 63%), *have* (266; 59%), *go* (246; 56%), *see* (210; 53%), *think* (208; 47%), *make* (101; 47%) and *come* (159; 45%). It is important to note that the classification of content words here are based on the impression of its meaning at face value and therefore, polysemous words are not considered until proven via close reading of the concordance lines.

Given these frequency lists, several observations can be made. Firstly, nouns that refer to temporal and spatial meanings are found salient, namely *time*, *night*, *way*, *day* and *everything*. These uses of deictic marker indicate time or moment of utterance (e.g., *time*, *night*, *day*) and distance or relative location of people and things (e.g., *way* and *everything*) (Yule, 2010), which may be descriptive of the way the message in the songs are written. As Tenbrink (2007) states, the use of temporal terms often emphasises connections between events, at the same time highlighting the significance of each event as well as communicating the writer or speaker's knowledge at the time. Spatial terms, on the other hand, are employed whenever a direction demands specification or when the writer or speaker wishes to communicate the distance from one entity to another as a way to emphasise contrast (Tenbrink, 2007). Meanwhile, *love*, *thing*, *baby*, *eye*, and *girl* could be grouped under person deixis as they describe a speaker or addressee, i.e., persons or subjects of the song as well as feelings related to them (Damayanti, 2022) and in turn, are also important components of song lyrics. Deixis therefore, can be used as signposts in song lyrics as they function to guide listeners when interpreting these songs and their meanings.

In terms of lexical verbs, the most frequently used ones shown in Table 4.2 can be sub-classified further into Cook's (1979) thematization of verbs, mainly basic verb types such as *do* and *make*, experiential types (*say*, *know*, *see*, *think*), benefactive types (*get*, *have*), and locative ones like *go* and *come*. These point to a higher use of experiential types of verbs where these particular verbs - i.e., *say*, *know*, *see* and *think* refer to how the songwriter expresses or shares her experiences with the audience and can be explored further by looking at context.

4.2 Collocation analysis

Following this, a collocation analysis was carried out for the top-ranked noun and lexical verb: *time* and *know*, to determine the strongest collocation for each word. Collocations, briefly, are a significant part of corpus analysis because by examining this, we can investigate patterns of co-occurrence between words observed in the corpus. As McEnery and Hardie rightfully point out, "collocation denotes the idea that important aspects of the meaning of a word [...] are not contained within the word itself [...], but rather subsist in the characteristic associations that the word [frequently] participates in" (2012: pp.122-123), hence it is important to not only examine keywords by themselves, but also its co-occurrence or collocation with other words. For this purpose, t-score was used as the statistical measure following Hunston (2002), where results will show the words that are significantly frequent to the node word (*time* and *know*, respectively), and threshold was increased from the default (3) to (5), as suggested by Baker (2016) to restrict the number of collocates that will yield a higher number of statistically significant collocates to be analysed. In addition, collocates should occur at least more than seven times in the corpus to be included in the analysis. The following sections describe these in detail.

4.2.1 Time

Figure 4.1 shows a visual representation of the collocation analysis using GraphColl (from #LancsBox) for the word 'time' based on the criteria described earlier. It can be seen that there are eight collocates of 'time', namely *the, me, I, you, and, a, this* and *all*. These suggest that the noun frequently occurs in various phrases such as 'the time', 'each time', 'every time' and 'this time' to name a few.

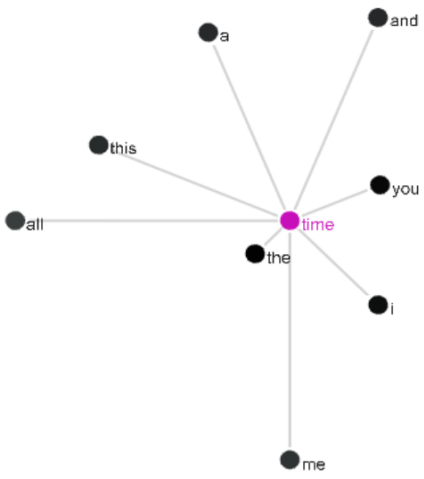


Figure 4.1 Collocates of 'time' based on t-score

More specifically, it can be seen that 'time' mostly collocates within these patterns or structures to function as temporal deixis, bringing the listener to a moment in her life where a significant event occurred that expanded on the meaning of the song. Examining the song Mastermind from 'Midnights', Swift uses time to direct her audience to key moments during the start of her relationship. It is possible that the song outlines her first introduction to Joe Alwyn, who then proceeded to become her partner at the time (Chow & Gutterman, 2022).

Once upon ***a time***, the planets and the fates and all the stars aligned
You and ***I*** ended up in the same room at ***the*** same ***time***
And the touch of a hand lit the fuse
Of a chain reaction of countermoves
To assess the equation of you
Checkmate, I couldn't lose
No one wanted to play with me as a little kid
So I've been scheming like a criminal ever since
To make them love me and make it seem effortless
This is ***the*** first ***time*** I've felt the need to confess
I laid the groundwork and then saw a wide smirk
On your face, you knew ***the*** entire ***time***
You knew that I'm a mastermind
And now you're mine
Yeah, all you did was smile
Cause I'm a mastermind.

While most songs follow this pattern, there are some exceptions in her discography. One song to note is Nothing New from 'Red (Taylor's Version)'. Instead of directing her audience to specific junctures, time is used to describe a change or progression from the song's present reality. Here, time is portrayed as an object or currency she will inevitably lose. This coincides with the song's themes; her fear of change and ageing as well as being irrelevant or replaced (Songfacts, 2023a).

And my cheeks are growing tired
From turning red and faking smiles
Are we only biding ***time*** 'til I lose your attention?

And someone else lights up the room?

People love an ingenue

I've had too much to drink tonight

How did I go from growing up to breaking down?

And I wake up in the middle of the night

It's like I can feel ***time*** moving

How can a person know everything at 18 but nothing at 22?

Will you still want me when I'm nothing new?

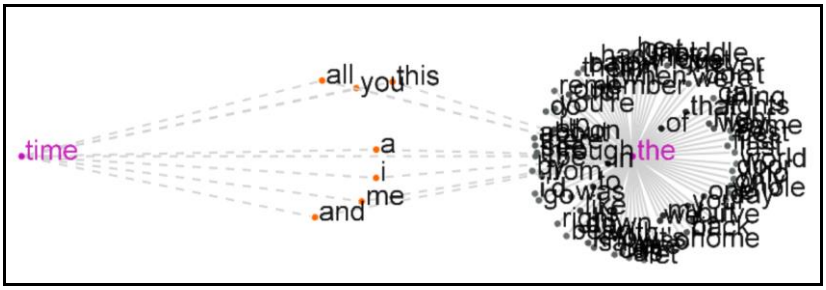


Figure 4.2 Collocation network for 'time' and 'the'

Further analysis of the frequent collocation 'time' + 'the' shows that this particular combination tends to occur with other words, namely *this*, *all*, *you*, *a*, *I*, *me*, and *and*. These point to overuse of certain phrases like 'all the time', 'the time I' and 'the time you' where they function to signify the state she is constantly in. A few occurrences include, *I regret you all the time* and *I go back to December all the time*. Its use can also highlight a key event that takes place in the songs like *Nothing made sense 'til the time I saw your face, this is the last time you* and *the first time you ever saw me cry*. These are function words that do not carry additional meaning to a sentence but will contribute to the syntax aspect of the sentence.

4.2.2 Know

Similarly, collocation analysis was carried out for the salient lexical verb 'know'. Figure 4.3 shows more collocates frequently co-occurring with 'know' (17 collocates) compared to 'time', which are *we*, *better*, *it*, *what*, *is*, *that*, *I'm*, *a*, *the*, *but*, *you*, *all*, *don't*, *and*, *to*, *I*, and *me*. In contrast to 'time', 'know' co-occurs more with the plural pronoun 'we' in 'we know', co-occurs with 'better' as an adjective to the lexical verb (*know better*), and collocates with other functional words like 'what' (signalling a question tag or relative clause), 'is' (be-verb to signal existentialism), 'that' (relative pronoun), contractions *I'm* and *don't*, conjunction 'but' and the preposition 'to'.

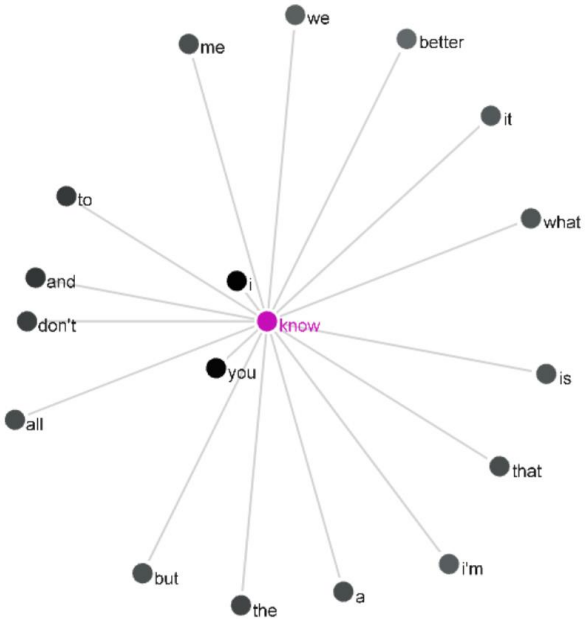


Figure 4.3 Collocation Analysis Result for 'know'

It can be seen that the first-person pronoun *I* had the most frequent collocation with 'know', and shared similar collocates that are *we, that, the, don't, it, me, a, you, to, is, what, but, I'm, all, better* and *and* (see Figure 4.4). This suggests that the phrase 'I know' strongly co-occurs with these types of words and may be repeated often in the lyrics that would contribute to its saliency. 'I know you' (e.g. *And I know you heard about me*; *You say you're fine I know you better than that*; *I know you were on my side even when I was wrong)*, 'I know what' (e.g. *Cause baby, I know what you know*; *I know what they all say but I ain't tryna play*) and 'I know better' (e.g. *I know better but you're still around*) are the examples of the lexical bundles that signifies her knowledge and awareness about a certain subject.

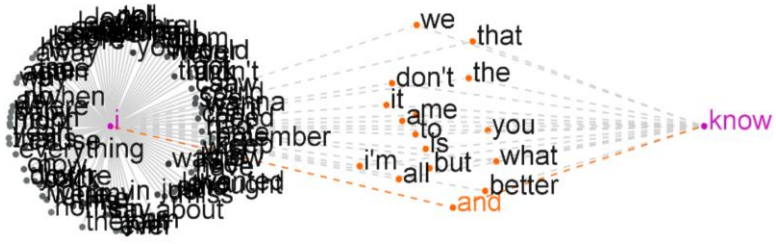


Figure 4.4 Collocation network for 'know' and 'I'

4.3 Lexical bundles

In identifying the occurrences of the lexical bundles in the corpus, an N-grams analysis of the corpus was carried out. As described earlier, key phrases or word combinations are generated automatically using the

'N-grams' function in LancsBox and only 3-grams (trigrams) were included in this analysis. Table 4.2 shows the top 10 most frequent three-word lexical bundles that can be found in the specialised corpus.

Table 4.2 Top 10 3-Word Lexical Bundles

	3-Word Lexical Bundles	**Freq.**	**Dis. (%)**
1	you and i	40	13.76
2	i don't wanna	27	5.82
3	i don't know	23	9.52
4	all i know	20	2.12
5	this is the	19	3.70
6	you know i	18	6.35
7	a better man	17	1.06
8	the last time	17	3.70
9	the middle of	16	5.29
10	i didn't know	16	4.76

Based on the above table, it can be seen that the phrase 'you and I' top the list, occurring 40 times across 13.76% of the songs in the corpus, followed by 'I don't wanna' (27; 5.82%), and 'I don't know' (23; 9.52%). However, it can be argued that 'I don't know' is more popular than 'I don't wanna' because it is distributed or occurs in a lot more songs. This is akin to other phrases that have higher dispersion scores (regardless of their frequency) like 'you know I' (18; 6.35%) and 'the middle of' (16; 5.29%). In turn, it is imperative to consider dispersion score or 'range' when describing key or word lists and for the purpose of investigating key lexical bundles in detail, only the first and third-ranked bundles (*you and I*, and *I don't know*, respectively) were analysed as they were both deemed to have higher frequency and dispersion scores combined. The concordance lines of these lexical bundles are shown in Figures 4.5 and 4.6, respectively.

Figure 4.5 presents instances of the frequent phrase/bundle 'you and I' in the corpus. Interestingly, this bundle can be seen to function in two main ways: 1) indicating the proximity or relationship between the singer-songwriter and another person (e.g., *can't stop thinkin bout* ***you and I***; *good times* ***you and I***), and 2) describing what/how singer-songwriter responds in relation to 'you' - typically denoted by the capital letter in 'And' or comma or question mark after 'you' (e.g., *since I have even heard from* <u>*you* ***A****nd I*</u> *should just tell you; Remember how I'd fly to* <u>*you****?*** *And I*</u> *can't talk to you; But I just miss* <u>*you****,*** *and I*</u> *just wish you*). Since song lyrics are poetic in nature, "[it] is a private expression of emotion by a single speaker" (Craven, 2020) and the lyrics are only meaningful given close reading of the concordance lines.

Search you and i **Occurrences** 40 (8.19) Texts 26/189 **▼ Corpus** Taylor Swift Corpus **▼ Context** 7

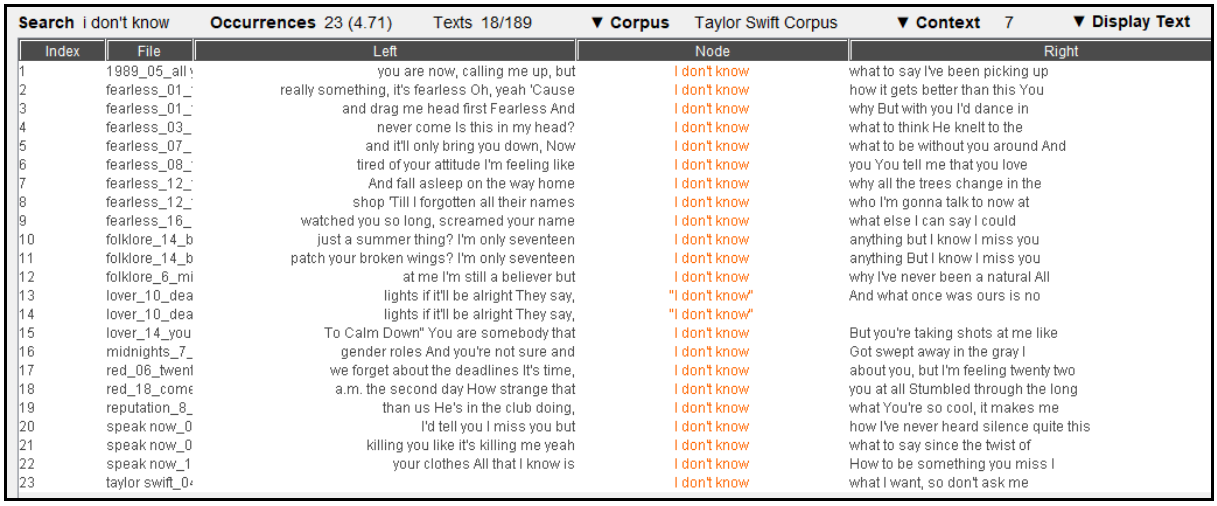
Index	File	Left	Node	Right
1	1989_03	while since I have even heard from	you And I	should just tell you to leave 'cause
2	1989_03	true but I Can't stop thinkin' 'bout	you and I"	I said "I've been there too a
3	1989_08	sad to think about the good times	You and I	Did you think we'd be fine? Still
4	1989_14	eye open at night We found wonderland	You and I	got lost in it And we pretended
5	1989_14	their mind But darling, we found wonderland	You and I	got lost in it And we pretended
6	1989_14	we both went mad We found wonderland	You and I	got lost in it And we pretended
7	evermor	flying 'til the bone crush Everybody wants	you And I	don't like a gold rush What must
8	fearless_	why are people always leaving? I think	you and I	should stay the same 'Cause I can't
9	lover_12	I won't try And I say to	you And I	hate to make this all about me
10	lover_13	separating us Remember how I'd fly to	you? And I	can't talk to you when you're like
11	lover_2_	don't wanna keep secrets just to keep	you And I	snuck in through the garden gate Every
12	lover_9_	hit the tunnel Sat on the roof,	you and I	You hold my hand on the street
13	midnight	the fates and all the stars aligned	You and I	ended up in the same room at
14	midnight	die? Years of tearing down our banners,	you and I	Living for the thrill of hitting you
15	midnight	once Believed in me And I felt	you and I	held you For a while Bet I
16	midnight	that I gave up And I lost	you And I	wake with your memory over me That's
17	red_07_	takes everything in me not to call	you And I	wish I could run to you And
18	red_07_	And I wish I could run to	you And I	hope you know that every time I
19	red_07_	me if I wanna try again with	you And I	almost do
20	red_08_	around again and say "Baby, I miss	you and I	swear I'm gonna change, trust me." Remember
21	red_21_	feel you again But I just miss	you, and I	just wish you were a better man
22	red_21_	of my hand But I just miss	you, and I	just wish you were a better man
23	red_21_	feel you again But I just miss	you, and I	just wish you were a better man
24	red_21_	of my hand But I just miss	you and I	just wish you were a better man
25	red_22_	I'm nothing new? But I just miss	you, and I	just wish you were a better man
26	red_22_	of my hand But I just miss	you and I	just wish you were a better man
27	red_25_	can't help who you fall for And	you and I	fell like an early spring snow But
28	reputatio	the memories, they will hold on to	you And I	will hold on to you Please don't
29	speak no		You and I	walk a fragile line I have known
30	speak no	not gone, you can't be gone, no	You and I	walk a fragile line I have known
31	speak no	kingdom lights shined just for me and	you And I	was screaming long live all the magic
32	taylor sw	thing, honey But I cried, cried for	you And I	know you wouldn't have told nobody if
33	taylor sw	if it do n't, stay beautiful If	you and I	are a story That never gets told
34	taylor sw	ran when you tried Just two kids,	you and I	Oh my, I was sixteen when suddenly
35	taylor sw	very front porch After all this time,	you and I	And I'll be eighty-seven; you'll be eighty-nine
36	taylor sw	every night And now you ask about	you and I	There's no you and I Remember what
37	taylor sw	ask about you and I There's no	you and I	Remember what you put me through I
38	taylor sw	had, oh baby I haven't thought about	you and I	There's no you and I And I
39	taylor sw	thought about you and I There's no	you and I	And I know Someday you will... Wake
40	taylor sw	stars In a field behind your yard	You and I	are painting pictures in the sky Sometimes

Figure 4.5 Concordance lines of 'you and I'

It was found that the occurrence of this lexical bundle was dispersed into all 10 albums excluding the 8th album – 'Folklore' (2020). The album 'Red' (2012) had the most occurrences compared to the remaining albums, with 10 occurrences in total. This may suggest that there are more references made about the singer-songwriter and another person compared to her depiction of songs in 'Folklore' (2020). As Kornhaber mentions, Folklore detracts from Swift's own experiences and instead shares fictional, historical and personal stories (2020). Therefore, the lack of the lexical bundle 'you and I' in this album is to be expected as she narrates the stories through song in third person, observing events unfolding as they happen.

In terms of the frequent bundle 'I don't know', Figure 4.6 presents instances where this bundle occurred in nine out of the ten albums in the corpus, except in the 9th album – 'Evermore' (2020). Generally, 'I don't know' is followed by WH-question tags like *what* (e.g., *but I don't know what to say*), *how* (e.g., *Cause I don't know how it gets better*), *why* (e.g., *and I don't know why*), and *who* (e.g., *I don't know who to talk to now*), which reflects an epistemic stance. This means that the singer-songwriter uses the bundle to depict her unknowing or uncertain position about something, thus further emphasising the degrees of uncertainty through use of WH-question tags. These specific instances mainly refer to what or how she should respond to something (I don't know what to say/ think/ I want) and her perception or opinion about something

(I don't know how it gets better/ to be something). Other occurrences of 'I don't know' include forms of scepticism like in *I'm only seventeen I don't know anything*, in relation to another person such as in *I don't know* (*about*) *you*, and one instance that occurs with a figurative meaning (*I don't know why all the trees change*).



Search i don't know **Occurrences** 23 (4.71) Texts 18/189 **▼ Corpus** Taylor Swift Corpus **▼ Context** 7 **▼ Display Text**

Index	File	Left	Node	Right
1	1989_05_all	you are now, calling me up, but	I don't know	what to say I've been picking up
2	fearless_01_	really something, it's fearless Oh, yeah 'Cause	I don't know	how it gets better than this You
3	fearless_01_	and drag me head first Fearless And	I don't know	why But with you I'd dance in
4	fearless_03_	never come Is this in my head?	I don't know	what to think He knelt to the
5	fearless_07_	and it'll only bring you down, Now	I don't know	what to be without you around And
6	fearless_08_	tired of your attitude I'm feeling like	I don't know	you You tell me that you love
7	fearless_12_	And fall asleep on the way home	I don't know	why all the trees change in the
8	fearless_12_	shop 'Till I forgotten all their names	I don't know	who I'm gonna talk to now at
9	fearless_16_	watched you so long, screamed your name	I don't know	what else I can say I could
10	folklore_14_b	just a summer thing? I'm only seventeen	I don't know	anything but I know I miss you
11	folklore_14_b	patch your broken wings? I'm only seventeen	I don't know	anything But I know I miss you
12	folklore_6_mi	at me I'm still a believer but	I don't know	why I've never been a natural All
13	lover_10_dea	lights if it'll be alright They say,	"I don't know"	And what once was ours is no
14	lover_10_dea	lights if it'll be alright They say,	"I don't know"	
15	lover_14_you	To Calm Down" You are somebody that	I don't know	But you're taking shots at me like
16	midnights_7_	gender roles And you're not sure and	I don't know	Got swept away in the gray I
17	red_06_twent	we forget about the deadlines It's time,	I don't know	about you, but I'm feeling twenty two
18	red_18_come	a.m. the second day How strange that	I don't know	you at all Stumbled through the long
19	reputation_8_	than us He's in the club doing,	I don't know	what You're so cool, it makes me
20	speak now_0	I'd tell you I miss you but	I don't know	how I've never heard silence quite this
21	speak now_0	killing you like it's killing me yeah	I don't know	what to say since the twist of
22	speak now_1	your clothes All that I know is	I don't know	How to be something you miss I
23	taylor swift_0		I don't know	what I want, so don't ask me

Figure 4.6 Concordance lines of 'I don't know'

The album that had the most occurrences of this phrase was 'Fearless' (2008) with 8 occurrences. The songs and the lines that contain the lexical bundle 'I don't know' in the album are shown in Table 4.3.1.

Table 4.3.1: Occurrences in 'Fearless'

Track	Song Name	Line
1	Fearless	Cause I don't know how it gets better than this
1	Fearless	And I don't know why but with you I'd dance in storm
3	Love Story	Is this in my head? I don't know what to think
7	Breathe	Now I don't know what to be without you around
8	Tell Me Why	I'm feeling like I don't know you
12	The Best Day	I don't know why all the trees change in the fall
12	The Best Day	I don't know who I'm gonna talk to now at school
16	Come in with The Rain	I don't know what else I can say

A closer look at these specific occurrences indicates that two occurrences of the bundle occurred in songs 'Fearless' and 'The Best Day'. To put things into context, 'Fearless' was described by Swift as "being about

the best first date that she has yet to have" (Songfacts, 2023b) so the description of being uncertain through use of the 'I don't know' bundle reflects her expression of excitement, particularly in the below verse:

And ***I don't know*** how it gets better than this
You take my hand and drag me head first
Fearless
And ***I don't know*** why
But with you I'd dance in a storm
In my best dress
Fearless

Meanwhile, 'The Best Day' is about a love letter to Taylor Swift's mother, Andrea Swift, who usually accompanies her daughter while she's on tour (Songfacts, 2023c). In this song, the two occurrences of 'I don't know' can be found in these two verses:

I don't know why all the trees change in the fall
But I know you're not scared of anything at all
Don't know if Snow White's house is near or far away
But I know I had the best day with you today

I don't know who I'm gonna talk to now at school
But I know I'm laughing
On the car ride home with you
Don't know how long it's gonna take to feel okay
But I know I had the best day with you today

5.0 Discussion

5.1 Linguistic items in Taylor Swift's song writing

As a reminder, the present study examined frequent lexical patterns and lexical bundles in Taylor Swift's song lyrics. Findings showed that the top 10 keywords in this corpus are function or closed class words such as pronouns 'you' and 'I'. This shows the personal and intimate nature of Taylor Swift's songs, placing the audience in the point of view of the singer and allowing them to associate better with the writer. Furthermore, while many critics have argued that the subject of Taylor Swift songs are mostly about love (Roberts, 2017), findings from this study showed that the word Swift uses the most is 'time' and its collocations act as

temporal deixis (e.g., *the same*/*first*/*entire time*). This indicates that her songs are written as a form of expression of her remembering or reminiscing about different times in her life. It can be argued that the overused word 'time' does revolve around aspects of her love life, but this may only prove that elements of love are rather expressed indirectly or implicitly.

The frequent uses of the verb 'know' on the other hand, indicated Swift's constant perceptions of others and herself regarding the topics mentioned in her songs, mainly *I know that I'm a handful baby*, *I know I'm better off alone*, *And I know I make the same mistakes every time* and *I know what they say*. In fact, the lexical verb 'know' was also found repeated in recurring strings of words or bundle, particularly in negated forms like *I don't know* and *I didn't know*, overgeneral statement as in *all I know*, and reference to audience in *you know I*. Further close readings of her lyrics that highlight these instances show how Swift perceives and reacts to the events, essentially expressing her views, cognitively. This further supports findings from Muhammad et al. (2021) that cognitive verbs like 'know' occur very frequently in contemporary song lyrics, reflecting the songwriter's thoughts and how he/she conceptualises the world (or experiences). These expressions depict Swift as knowing what the other subject or person in the song is thinking (e.g., *And I know you heard about me*,). This further communicates Swift's understanding of the situation she is in. However, it can portray other aspects of her character, where she perceives her assumptions as the absolute reality (e.g., *I know what they all say but I ain't tryna play*). It can also be interpreted as her desire to find out more about the subject, leading us to assume a curiosity or perceptiveness in her nature (as in the use of 'I don't know').

5.2 Most used lexical bundles

The results showed that Swift's most used lexical bundles are 'you and I' as well as 'I don't know', further highlighting pronouns (*I* and *you*) and 'know' as salient features in her song writing. Utilising personal pronouns like 'you' or 'I' is essential in face-to-face interaction, allowing users to define or divulge interpersonal relationships between or amongst people participating in the interaction (Kuo, 1999). This further emphasises the personal nature of Swift's songs as well as allowing listeners to relate the pronouns with their own experiences. However, it must be pointed out that the repeated string of words 'you and I' did not necessarily mean the combination of pronouns with conjunction 'and' - such as in the example: *can't stop thinkin bout* ***you and I***, rather most instances were seen to be extractions of either a previous line that talks about 'you' followed by the continuing line beginning with 'and I' (e.g., *But I just miss you, and I just wish you*). One way to explain this is that song lyrics are rather poetic in nature and that according to Craven (2020), "[a] lyric poem is short, highly musical verse that conveys powerful feelings" and in turn, meanings would only be discerned by close readings of the concordance lines.

Another frequently used lexical bundle is 'I don't know', which communicates the attributes or feelings of the writer to the audience or epistemic stance. The expression of not knowing can be interpreted as naivete or innocence (or portraying oneself to be innocent) to the audience when looked at from a writer's perspective. It can also communicate the writer's awareness of her own uncertainty and the writer acknowledges it, showing maturity. However, taking into consideration how songs allow listeners to place themselves in the writer's seat, this appeals to the listener's view of themselves. As Djikic (2011) has asserted, lyrics are mostly limited to personal content, allowing listeners to identify and contrast themselves while further emphasising ideas they have about their own personalities.

In terms of Swift's frequent use of these bundles in the corpus, we could posit some insight into its uses in her selected albums. The lexical bundle 'you and I' was not found in the album 'Folklore', which coincides with the album's direction that centred around the stories of others as opposed to her previous songs that were autobiographical in nature (Mapes, 2020). Meanwhile, the frequent bundle 'I don't know' which appeared more in her earlier work as compared to her more recent work, gives us a glimpse of how she viewed herself when she first started out in the past and how she is in the present - mainly of her lack of experience such as shown in *I'm only seventeen I don't know anything*. In contrast, in her more recent work, she utilises it differently, using it to express that she does not know what to do next, and therefore expresses shared experiences with listeners like in *I don't know* (*about*) *you*. This also further increased the appeal of her work and explains its popularity amongst younger audiences, while also showing her maturity throughout the years in her expressions within the lyrics.

6.0 Conclusion

The availability of corpus research on a body of work from a single songwriter is rather scant. However, the benefits of such research present readers (or in this case, listeners) to the creative writing styles and motivations of a songwriter. This paper reports on Taylor Swift's style of writing and presents her discography as a suitable choice for corpus analysis on song lyrics. The results show that a songwriter's lyrics are forms of the artist's communication of their feelings, emotions as well as hopes and dreams whether written intentionally or unintentionally. This was found throughout salient occurrences of topical words that depict Swift's choice of lyrics, namely in terms of the frequently used lexicon 'time' and functional word 'know'. Further analysis of these salient items revealed that she mostly writes songs in relation to specific time periods and what or how she knows or perceives them. In other words, Taylor Swift's songs are described as a piece of storytelling, mainly about her life or (personal) journey. However, the present study only examined the top most frequent closed and open class words in her songs, their frequent collocations as well as frequently used lexical bundles. Future studies could investigate other properties of the artist's writing such as metaphorical or figurative use of language. In addition, similar corpus approaches or comparative type ones

can be carried out to a wider set of song lyrics, not limited to one songwriter. This is because music and its lyrics are integral to human expressions, understanding why songs are important through analysis of lyrics would enable listeners to better appreciate them.

References

Atar, C., & Erdem, C. (2019). The advantages and disadvantages of corpus linguistics and conversation analysis in second language studies. *Proceedings of IX Scientific and Practical Internet Conference of Young Scientists and Students*, 140-146. https://www.researchgate.net/publication/337858444

Baker, P. (2016). The shapes of collocation. *International Journal of Corpus Linguistics 21*(2), 139-164. https://www.academia.edu/28797223/Baker_P_2016_The_shapes_of_collocation_International_Journal_of_Corpus_Linguistics_21_2_139_164

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Boonjoon, N. (2018). A corpus-based study of Taylor Swift's songs: the linguistic characteristics and the shift in genre. https://www.researchgate.net/publication/328365180_A_Corpus-based_study_of_Taylor_Swift's_songs_the_linguistic_characteristics_and_the_shift_in_genre

Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, *6*, 1-9.

Choi, K., & Downie, J. S. (2019). A Trend Analysis on Concreteness of Popular Song Lyrics. In 6th International Conference on Digital Libraries for Musicology (DLfM '19). Association for Computing Machinery, New York, NY, USA, 43-52. https://doi.org/10.1145/3358664.3358673

Chow, A. R., & Gutterman, A. (2022, October 21). A Close Read of Every Line of Taylor Swift's "Mastermind." *Time*. https://time.com/6223855/taylor-swift-mastermind-lyrics-explained/

Climent, S., & Coll-Florit, M. (2020). All you need is love: metaphors of love in 1946–2016 Billboard year-end number-one songs. *Text & Talk*, *41*(4), 469–491. https://doi.org/10.1515/text-2019-0209

Cook, W. A. (1979). *Case Grammar: Development of the Matrix Model (1979-1978)*, Georgetown University Press.

Craven, J. (2021, February 17). *Lyric Poetry: Expressing Emotion Through Verse*. ThoughtCo. https://www.thoughtco.com/lyric-poem-definition-examples-4580236

Damayanti, N. K. L. P. (2023). A deixis analysis of song lyrics in "Here's Your Perfect" by Jammie Miller. *Jurnal Penelitian Mahasiswa Indonesia*, *3*(1), 148-152. https://jurnal.stkipahsingaraja.ac.id/index.php/jpmi/article/view/470

Djikic, M. (2011). The effect of music and lyrics on personality. *Psychology of Aesthetics Creativity and the Arts*, *5*(3), 237–240. https://doi.org/10.1037/a0022313

Eiter, A. (2017). *'Haters gonna hate': A corpus linguistic analysis of the use of non-standard English in pop songs* [BA Thesis]. University of Innsbruck. https://www.academia.edu/33520286/Haters_gonna_Hate_A_Corpus_Linguistic_Analysis_of_the_Use_of_Non_Standard_English_in_Pop_Songs/

Ghane, Z., & Farahani, M. V. (2021). Doing linguistics with a corpus: Methodological considerations for the everyday user. Egbert Jesse, Larsson Tove and Biber Douglas, Cambridge: Cambridge University Press, 2020. *International Journal of Applied Linguistics*, *31*(2), 321–323. https://doi.org/10.1111/ijal.12373

Gries, S. T. (2020). Corpus linguistics and the law: extending the field from a statistical perspective. *Brook. L. Rev.*, *86*, 321. https://brooklynworks.brooklaw.edu/blr/vol86/iss2/2

Hadian, M. (2015). The use of song lyrics in teaching listening. *Journal of English and Education*, *3*(1), 96-105.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press, Cambridge.

Knees, P., Schedl, M., & Widmer, G. (2005, October). Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics. In *ISMIR* (pp. 564-569).

Krasse, L. (2019). *A corpus linguistic study of the female role in popular music lyrics* [BA thesis, Malmö University]. Malmö University Publications. https://mau.diva-portal.org/smash/record.jsf?pid=diva2%3A1481063&dswid=-9806

Kreyer, R. (2015). "Funky fresh dressed to impress." *International Journal of Corpus Linguistics*, *20*(2), 174–204. https://doi.org/10.1075/ijcl.20.2.02kre

Kreyer, R., & Mukherjee, J. (2007). The style of pop song lyrics: A corpus-linguistic pilot study. *Anglia*, *125*(1), 31-58. https://doi.org/10.1515/angl.2007.31

Kuo, C.-H. (1999). The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles. *English for Specific Purposes*, *18*(2), 121–138. https://doi.org/10.1016/S0889-4906(97)00058-6

Lee, D. Y. W. (2008). Corpora and discourse analysis: new ways of doing old things. In V. K. Bhatia, J. Flowerdew & R. Jones (Eds.), *Advances in discourse studies* (pp. 86–99). Routledge.

Logan, B., Kositsky, A., & Moreno, P. (2004, June). Semantic analysis of song lyrics. In *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)* (Vol. 2, pp. 827-830). IEEE.

Lyon, L., R. (2019). *Shifting Personas: A Case Study of Taylor Swift* [MA Thesis]. University of Kentucky.

Mahlberg, M. (2010). Corpus linguistics and the study of nineteenth-century fiction. *Journal of Victorian Culture*, *15*(2), 292-298.

Mapes, J. (2020, July 27). *Folklore, Taylor Swift*. Pitchfork. https://pitchfork.com/reviews/albums/taylor-swift-folklore

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.

McEnery, T., & Hardie, A. (2013). The History of Corpus Linguistics. In *Oxford University Press eBooks* (pp. 727–745). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199585847.013.0034

Muhammad, M. M., Goyak, F., Zaini, M. F., & Abdelkhalek Ibrahim, W. M. (2022). Diachronic Analysis of the Profane Words in English Song Lyrics:: A Computational Linguistics Perspective. *Malaysian Journal of Music, 11*(1), 14–32. https://doi.org/10.37134//mjm.vol11.1.2.2022

Nishina, Y. (2017). A study of pop songs based on the Billboard Corpus. *International Journal of Language and Linguistics, 4*(2), 125–134. http://ijllnet.com/journals/Vol_4_No_2_June_2017/16.pdf

Oliveira, H. G. (2015). Tra-la-lyrics 2.0: Automatic generation of song lyrics on a semantic domain. *Journal of Artificial General Intelligence*, *6*(1), 87-110.

Racette, A., & Peretz, I. (2007). Learning lyrics: to sing or not to sing?. *Memory & cognition*, *35*(2), 242-253.

Roberts, A. (2017, October 21). *Taylor Swift Is Already Getting Shamed For Writing Another Love Song & This Narrative Is Getting Old*. Bustle. https://tinyurl.com/yc8bjen7

Setiawati, W., & Maryani, M. (2018). An analysis of figurative language in Taylor Swift's song lyrics. *PROJECT (Professional Journal of English Education), 1*(3), 261. https://doi.org/10.22460/project.v1i3.p261-268

Simanjuntak, M. B., Zuriyati, Z., & Lustyantie, N. (2022). Metamorphic analysis in song lyrics Batak Toba "Jujung Goarhi Amang" (study of portray of parents). *PROJECT (Professional Journal of English Education), 5*(1), 236-243. https://journal.ikipsiliwangi.ac.id/index.php/project/article/view/10000

Songfacts. (2023b). *Fearless by Taylor Swift*. https://tinyurl.com/yh6mautz

Songfacts. (2023c). *The Best Day by Taylor Swift*. https://tinyurl.com/mwp768zb

Syafar, D. N., & Asty, H. (2022). Homonymy analysis in Taylor Swift's Songs. *Jurnal Ilmiah Langue and Parole*, *6*(1), 7–12. https://doi.org/10.36057/jilp.v6i1.546

Tenbrink, T. (2007). *Space, Time, and the Use of Language: An Investigation of Relationships*. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110198829

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.

Wolk, D. (2012, November 13). One Direction's Songwriters: They're What Make the Boy Band Beautiful. TIME.com. https://entertainment.time.com/2012/11/13/one-directions-songwriters-theyre-what-make-the-boy-band-beautiful/

Wulff, S. (2015). What learner corpus research can contribute to multilingualism research. *International Journal of Bilingualism, 21*(6), 734-753. https://doi.org/10.1177/1367006915608970

Yeh, A., Hung, T., & Candidate, M. A. (2022). The stages of love, songs, and a band: A corpus discourse analysis of One Direction's pop albums. *Advances in Language and Literary Studies*, *13*(1), 70-81. https://doi.org/10.7575/aiac.alls.v.13n.1.p.70

Yule, G. (2010). *The Study of Language (4th ed.)*. Cambridge: Cambridge University Press.

Yusnitasari, V., Wangi, W., & Sugianto, Y. (2022). An analysis of figurative language on the song lyrics "You are My Sunshine" by Anne Murray. *LUNAR: Journal of Language and Art*, *6*(1), 309-322.

Appendix

Taylor Swift Corpus (TSC)

1989_01_welcome to new york
1989_02_blank space
1989_03_style
1989_04_out of the woods
1989_05_all you had to do was stay
1989_06_shake it off
1989_07_i wish you would
1989_08_bad blood
1989_09_wildest dreams
1989_10_how you get the girl
1989_11_i know places
1989_12_this love
1989_13_clean
1989_14_wonderland
1989_15_you are in love
1989_16_new romantics
evermore_1_willow
evermore_2_champagne problems
evermore_3_gold rush
evermore_4_tis the damn season
evermore_5_tolerate it
evermore_6_no body no crime
evermore_7_happiness
evermore_8_dorothea
evermore_9_coney island
evermore_10_ivy
evermore_12_long story short
evermore_13_marjorie
evermore_14_closure
evermore_15_evermore
evermore_16_right where you left me
evermore_17_its time to go
evermore_111_cowboy like me
fearless_01_fearless
fearless_02_fifteen
fearless_03_love story
fearless_04_hey stephen
fearless_05_white horse
fearless_06_you belong with me
fearless_07_breathe
fearless_08_tell me why
fearless_09_you_re not sorry
fearless_10_the way i loved you
fearless_11_forever and always
fearless_12_the best day
fearless_13_change
fearless_14_jump then fall
fearless_15_untouchable
fearless_16_come in with the rain
fearless_17_superstar
fearless_18_the other side of the door
fearless_tv_19_today was a fairytale
fearless_tv_20_you all over me
fearless_tv_21_mr perfectly fine
fearless_tv_22_we were happy
fearless_tv_23_thats when
fearless_tv_24_dont you
fearless_tv_25_bye bye baby
folklore_1_the1
folklore_2_cardigan
folklore_3_the great american dynasty
folklore_4_exile
folklore_5_my tears ricochet
folklore_6_mirrorball
folklore_7_seven
folklore_8_august
folklore_9_this is me trying
folklore_10_illicit affairs
folklore_11_invisible string
folklore_12_mad woman
folklore_13_epiphany
folklore_14_betty
folklore_15_peace
folklore_16_hoax
folklore_17_the lakes
lover_1_i forgot that you existed
lover_2_cruel summer
lover_3_lover
lover_4_the man
lover_5_the archer
lover_6_i think he knows
lover_7_miss americana & the heartbreak prince
lover_8_paper rings
lover_9_cornelia street
lover_10_death by a thousand cuts
lover_11_london boy
lover_12_soon youll get better
lover_13_false god
lover_14_you need to calm down
lover_15_afterglow
lover_16_me
lover_17_its nice to have a friend
lover_18_daylight
midnights_1_lavender haze
midnights_2_maroon
midnights_3_anti-hero
midnights_4_snow on the beach
midnights_5_youre on your own kid
midnights_6_midnight rain
midnights_7_question...
midnights_8_vigilante shit
midnights_9_bejeweled
midnights_10_labyrinth
midnights_11_karma
midnights_12_sweet nothing
midnights_13_mastermind
midnights_14_the great war
midnights_15_bigger than the whole sky
midnights_16_paris
midnights_17_high infidelity
midnights_18_glitch
midnights_19_wouldve couldve shouldve
midnights_20_dear reader
midnights_21_hits different
red_01_state of grace
red_02_red
red_03_treacherous
red_04_i knew you were trouble
red_05_all too well 10min version
red_06_twenty two
red_07_i almost do
red_08_we are never ever getting back together
red_09_stay stay stay
red_10_the last time
red_11_holy ground
red_12_sad beautiful tragic
red_13_the lucky one
red_14_everything has changed
red_15_starlight
red_16_begin again
red_17_the moment i knew
red_18_come back be here
red_19_girl at home
red_20_ronan
red_21_better man
red_22_nothing new
red_23_babe
red_24_message in a bottle
red_25_i bet you think about me
red_26_forever winter
red_27_run
red_28_the very first night
reputation_1_...ready for it
reputation_2_endgame
reputation_3_i did something bad
reputation_4_dont blame me
reputation_5_delicate
reputation_6_look what you made me do
reputation_7_so it goes...
reputation_8_gorgeous
reputation_9_getaway car
reputation_10_king of my heart
reputation_11_dancing with our hands tied
reputation_12_dress
reputation_13_this is why we cant have nice thi...
reputation_14_call it what you want
reputation_15_new years day
speak now_01_mine
speak now_02_sparks fly
speak now_03_back to december
speak now_04_speak now
speak now_05_dear john
speak now_06_mean
speak now_07_the story of us
speak now_08_never grow up
speak now_09_enchanted
speak now_10_better than revenge
speak now_11_innocent