

A Multi-Dimensional Analysis of Western Boy Band and Western Girl Group Song Lyrics

Amany Awad Shaaban Hussein (*)

1. Introduction

The field of Corpus Linguistics has undergone a remarkable evolution since its inception in the mid-20th century. The foundations of Corpus Linguistics can be traced back to the 1950s and 1960s when researchers started to compile and analyze large collections of texts to study language patterns. Initially emerging as an idea to study language through the analysis of larger collections of texts, Corpus Linguistics took a significant leap forward with the advent of computational tools in the 1960s and 1970s. The development of computers enabled linguists to compile, store, and analyze vast corpora of texts efficiently, giving rise to a data-reliant approach to studying language. These corpora became valuable resources for linguistic inquiry, allowing scholars to explore syntax, semantics, and pragmatics, while influencing the development of new linguistic theories.

One of the most influential methodologies within Corpus Linguistics is the Multi-Dimensional Analysis (MDA) approach, introduced by Douglas Biber in 1988. MDA represents a shift from traditional linguistic models by focusing on actual language use across different genres and contexts. This approach employs statistical methods to identify patterns of linguistic variation, providing insights into how language varies across genres and influencing research in areas such as sociolinguistics, stylistics, and discourse analysis. Studies

(*) This paper is part of Ph.D thesis entitled: "A Corpus-Linguistic Analysis of Western Boy Band and Western Girl Group Song Lyrics". Supervised by Prof. Bahaa-eddin M. Mazid- Faculty of Al-Alsun (Languages) - Sohag University & Dr. Ismail Abdelghany Ahmed - Faculty of Arts- Sohag University.

using MDA have explored linguistic variation in numerous genres, from academic writing to web-based texts, and have even generated new dimensions of analysis.

In recent years, the application of corpus-linguistic methods to song lyrics has gained attention among researchers, particularly as digital platforms have revolutionized music consumption and accessibility. Researchers have begun to systematically analyze the linguistic features of song lyrics across various music genres, highlighting the unique characteristics of this artistic genre (e.g. Bértoli-Dutra, 2014; Ma, 2017; Schneider and Miethaner, 2006; Kreyer and Mukherjee, 2007; Bridle, 2011; Falk, 2012).

Building on these studies, the present research utilizes Biber's (1988) MDA approach to investigate linguistic patterns and dimensions of variation in a corpus of pop song lyrics performed by Western boy bands and girl groups. The objective of this research is to analyze over 52 lexicogrammatical features within the Western Boy Band and Western Girl Group Pop Song Lyrics (WBB&WGGPSL) Corpus, aligning them with the characteristics of 23 written and spoken English genres examined in Biber's (1988) study. The study aims to pinpoint the location of this corpus along Biber's first three dimensions of variation, identifying similarities and differences with the other genres.

The significance of the present research lies in its ability to reinforce the notion that genre variation in language is both pervasive and systematic. By applying Biber's MDA model to a new dataset, the research not only validates the model's applicability but also extends its framework, particularly in Dimension 1 (Involved vs. Informational Production). This study introduces three novel linguistic features—interjections, base form verbs (used in imperative sentences), and first-person subject omission—that enhance the understanding of language use in songwriting. These features contribute to the

involved aspect of the lyrics by fostering emotional immediacy and brevity.

2. Literature Review

The analysis of song lyrics through corpus-based methodologies has provided valuable insights into the linguistic characteristics of various musical genres. Biber's (1988) MDA Model has been a significant tool in some of these studies. For example, Bértoli-Dutra (2014) applied MDA to a comprehensive corpus of North American and British pop songs spanning from 1940 to 2009, which included 6,290 individual songs from 32 artists across 16 genres. This study employed automatic part-of-speech tagging and semantic field analysis, as well as the identification of multi-word units through comparisons with a Google n-gram corpus. Her analysis revealed two primary sets of dimensions—lexico-grammar and semantics—that mirrored Biber's dimensions. Notably, her study found that pop lyrics exhibited key linguistic features similar to Biber's dimensions of *Overt Expression of Persuasion* and *Involved versus Informational Production*, while also uncovering semantic themes related to personal experiences and the world of music itself.

Ma (2017) extended Biber's MDA Model to analyze a corpus of 770 top-ranking modern English songs from the US Billboard charts, covering the years 1940 to 2016. Her study discovered that these songs exhibited high levels of listener involvement, moderate narrative concerns, and overt persuasiveness, with relatively low levels of online elaboration. By comparing these dimensions with 23 spoken and written genres, Ma identified distinctive features such as increased verbalization over time and genre-specific language trends. Her findings also highlighted that Rock music, in particular, showed a strong focus on current events and a shift in emphasis from the artist to the object of their lyrics.

Other studies have conducted a corpus-linguistic analysis of song lyrics using different theoretical frameworks. To begin

with, Schneider and Miethaner (2006) introduced the Blues Lyrics corpus (BLUR), which includes over 8,000 blues lyrics from the early 20th century. Their study utilized this corpus to investigate African American Vernacular English (AAVE) and identified unique syntactic constructions in blues lyrics that had previously been unrecognized. By comparing these features with other electronic corpora, Schneider and Miethaner (2006) demonstrated the value of blues lyrics in tracing historical language patterns and linguistic innovations.

Kreyer and Mukherjee (2007), on the other hand, focused on stylistic variation within pop song lyrics using the Giessen-Bonn Corpus of Popular Music (GBoP). Their study aimed to identify general style markers and variations within the genre by analyzing vocabulary and lexico-grammatical features. They also examined metaphorical usage in pop lyrics, providing a comprehensive view of the stylistic and thematic elements characteristic of this genre.

Moreover, Bridle (2011) utilized Wmatrix to analyze blues lyrics from 35 male artists, focusing on linguistic features across different historical periods. The study examined key domains, words, and parts of speech to understand how blues lyrics reflect social and political conditions of their times. By investigating the work of Robert Johnson and comparing it with broader blues trends, Bridle aimed to elucidate the role of blues lyrics in shaping African American identity and linguistic culture.

Furthermore, Falk (2012) explored the language of rock lyrics using a corpus of approximately 53,000 words. Her study sought to identify genre-specific features and compare rock lyrics with general language usage and other musical genres. The results indicated that rock lyrics share some characteristics with other genres but also exhibit unique features, providing insights into how rock music mirrors or deviates from broader linguistic trends.

3. Methodology

The Corpus used in the present study, named the *Western Boy Band and Western Girl Group Pop Song Lyrics Corpus* (WBB&WGGPSL Corpus), was compiled specifically for this research project. The initial phase in the data collection process involved the careful selection of Western boy bands and girl groups whose songs fell within the pop genre and who achieved fame during the musical time periods of the 1990s and 2000s. This procedure resulted in the selection of five boy bands and five girl groups. These five boy bands and girl groups are: The Backstreet Boys, Westlife, *NSYNC, Boyzone, Blue, All Saints, Atomic Kitten, B*Witched, The Pussycat Dolls, and Spice Girls.

The following two steps were to decide which song lyrics of each boy band and each girl group to include in the corpus, and to collect the lyrics of these songs. The subsequent procedure involved gathering written lyrics from an online source of song lyrics and organizing them systematically. For this purpose, the website AZLyrics (<https://www.azlyrics.com/>), which provides a vast database of song lyrics from various genres and artists, was chosen as the online source used for collecting the song lyrics corpus.

In order to increase the probability of obtaining a representative sample and minimize selection bias that may arise if songs were chosen based on popularity or personal preferences, random selection was employed in the choice of which songs to include in the corpus. Microsoft Excel was utilized to facilitate the random selection of song lyrics from the five boy bands and the five girl groups.

The following step incorporated the manual extraction of the lyrics of the chosen songs from AZLyrics. The next step was to verify the accuracy of the lyrics obtained from AZLyrics against the recorded versions of the songs in order to check for any discrepancies. This involved engaging in a comprehensive process of listening to each of the 300 songs

while simultaneously reading the lyrics. The corpus of the present study comprises 300 song lyrics: 150 sung by boy bands and 150 by girl groups; each band and group contributes 30 song lyrics to the corpus.

Two computer-based instruments of data analysis and visualization were used in the present study. These are: The Multi-Dimensional Analysis Tagger (MAT) (Nini, 2019), which was used in tagging and analyzing 52 features of Biber's (1988) study in addition to the Stanford Tagger's additional linguistic features; and Microsoft Excel, which was used in organizing, exploring, and visualizing the collected data and the results of the statistical analysis performed by MAT (Nini, 2019).

Following the compilation and structuring of the WBB&WGGPSL Corpus through organizing and formatting the corpus data to ensure compatibility with the software program, the corpus was tagged and annotated using the Multi-Dimensional Analysis Tagger (MAT) (Nini, 2019). Since the MD Analysis carried out in the present study is based on identifying rates of occurrence for a wide range of linguistic features which were generated using automatically tagged corpora, it was crucial to investigate the accuracy of the POS tagger used in Nini's (2019) MAT software program in the annotation of the WBB&WGGPSL Corpus. In order to enhance the accuracy of linguistic annotations, a meticulous manual revision was undertaken. The manual editing process involved examining the 300 MAT tagged texts to identify areas where the automatic tagging produced errors or inaccuracies. Observed errors in the automatic tagging were identified and corrected manually.

Once the corpus was fully tagged and the verification process was completed, the features were counted and analyzed against Biber's (1988) dimensions of variation using MAT (Nini, 2019). The results of the present corpus were compared to those of Biber's (1988) 23 written and spoken genres. These

genres are: Press reportage; Editorials; Press reviews; Religion; Skills and hobbies; Popular lore; Biographies; Official documents; Academic prose; General fiction; Mystery fiction; Science fiction, Adventure fiction; Romantic fiction; Humor; Personal letters; Professional letters; Face-to-face conversations; Telephone conversations; Public conversations, debates, and interviews; Broadcasts; Spontaneous speeches; and Planned speeches.

4. Findings

4.1 Dimension 1: Involved Versus Informational Production

Dimension 1: involved versus information production, comprises 34 linguistic features, many of which are fundamental to the existence of variation among texts; as Biber (1988) stresses that this dimension is a “basic dimension of variation among spoken and written texts in English” (p. 104). High scores on this dimension mean that the texts are characterized by high frequencies of private verbs, that-deletions, contractions, present tense verbs, second person pronouns, etc., and low frequencies of nouns and prepositions, low word lengths and low type to tokens ratio. On the other hand, low scores on this dimension mean that the texts have the opposite linguistic characteristics: very high frequencies of nouns, prepositional phrases, etc., plus low frequencies of private verbs, contractions, etc. Dimension 1 pertains to interactive involvement, which measures the degree of engagement, interaction, or connection between the communicator and the audience as opposed to a more detached, informational style. High scores on this dimension represent more involvement with listeners or readers and assess the level of interpersonal communication within the text; on the other hand, more detachment and informational content is associated with low scores on this dimension.

WBB&WGGPSL Corpus has a Dimension 1 score of 36.67 (Figure 1) which is a high score among Biber’s (1988)

genres. This result suggests that WBB&WGG pop song lyrics exhibit a notable level of interactive involvement. This involved, non-informational focus of boy band and girl group pop song lyrics is due to the existence of linguistic features that are characterized as being “verbal, interactional, affective, fragmented, reduced in form, and generalized in content” (Biber, 1988, p. 105).

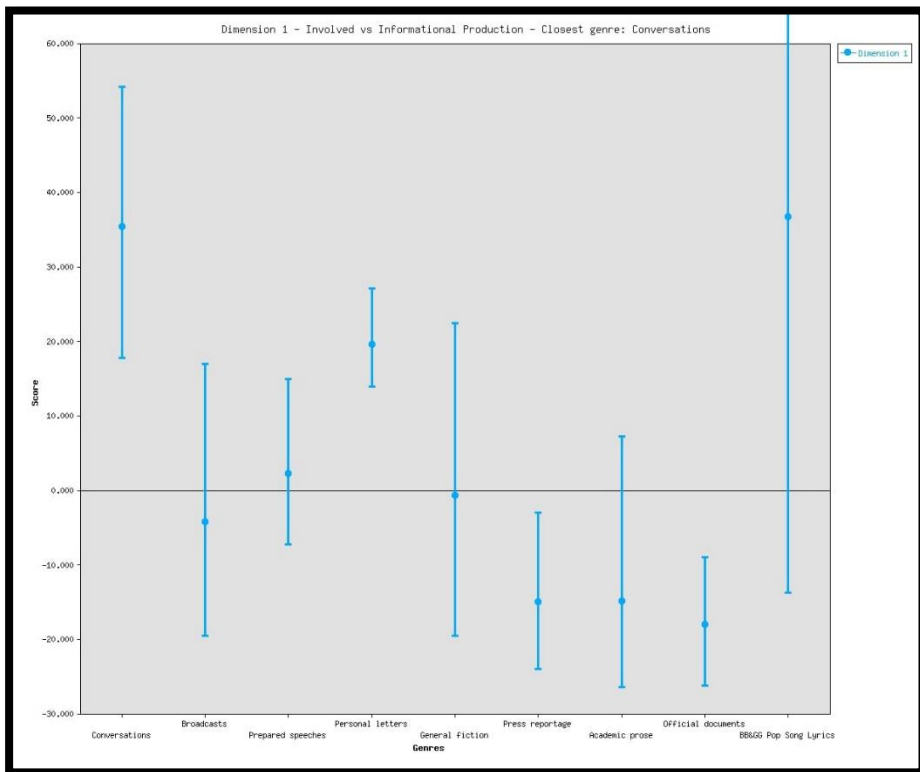


Figure 1. MAT’s visualization of the WBB&WGGPSL Corpus along Dimension 1

The lyrics of the corpus under investigation are characterized by features that actively engage the audience, such as direct address which occurs when the songwriter (or singer) directly speaks to the listener or a specific person within the lyrics. It involves a direct communication or engagement with the

audience, using first person pronouns (Z-Score = 3.13) and second person pronouns (Z-Score = 4.18). This technique creates a sense of immediacy and personal connection, as if the artist is speaking directly to the listener, which makes the lyrics more personal, relatable, and emotionally charged.

Private verbs (Z-Score = 1.72) and present tense forms (Z-Score = 0.94) are also among the features which are frequently used in WBB&WGG pop song lyrics. They indicate a verbal, interactive, and involved style. Private verbs, also known as mental or psychological verbs, are verbs that convey internal mental processes, emotions, or states of being. These verbs often provide insight into the thoughts, feelings, or perceptions of the communicator as they are used for “the overt expression of private attitudes” (Biber, 1988, p. 105). The frequent use of private verbs in WBB&WGG pop song lyrics contributes to the expression of the artist’s personal experiences and emotional states, which enhances the lyrical content by offering glimpses into the artist’s inner world (e.g. *know, think, feel, mean, pretend, believe, see, guess, fear, hear, remember, dream, realize, learn*). Moreover, the use of present tense forms contributes to a sense of immediacy and relevance in the communication as the present tense “refers to actions occurring in the immediate context of interaction” (Biber, 1988, p. 105). Thus, the frequent use of the present tense in WBB&WGG pop song lyrics places the events, emotions, or experiences in the current moment, which enhances the feeling of a dynamic and immediate connection between the artist and the audience. This temporal choice aligns with a contemporary and engaged style, as if the artist is sharing their thoughts and emotions in a real-time dialogue. Hence, the utilization of private verbs and present tense forms in WBB&WGG pop song lyrics is indicative of a communicative strategy that serves as a vehicle for conveying authenticity and immediacy and stresses the expression of subjective experiences and emotions in the present moment.

Other frequently used linguistic devices and features that enhance the sense of connection between the artist and the audience and mark a reduced surface form are that-deletions (Z-Score = 1.86); contractions (Z-Score = 3.5); pronoun it (Z-Score = 1.4); demonstrative pronouns, which are used less frequently than other linguistic features indicating involvement (Z-Score = - 0.22); and indefinite pronouns (Z-Score = 0.81). The reduced surface form of boy band and girl group pop song lyrics creates a more intimate, relatable, and conversational tone. That-deletions involve omitting the word ‘that’ when it is not essential for understanding an utterance. This creates a more conversational and streamlined style, making the lyrics more immediate and personal (e.g. *‘I think you lost it’*, *‘I swear I tell the truth’*, *‘I knew you changed the game’*). Contractions, on the other hand, refer to omitting letters, usually vowels, from the middle of words. They may be standard or colloquial. Colloquial contractions communicate emotions of affection; and, therefore, they establish tone. Colloquial contractions are reduced written forms that reflect informal pronunciation: reduction (e.g. *gonna*, *wanna*, *kinda*, *gotta*, *‘cause*, *cuz*, *lil*, *ma*, *Imma*); silent *h* (e.g. *av*, *v*, *s*); and assimilation (e.g. *gimme*, *lemme*). The frequent use of contractions in WBB&WGG pop song lyrics, most of which are informal, communicates sincere affection; as they add an informal and colloquial tone as “a consequence of fast and easy production” (Biber 1988, p. 243), contributing to a sense of rapport, intimacy, and closeness between the artist and the audience. Moreover, the use of the pronoun ‘it’ creates a sense of shared experience. In WBB&WGG pop song lyrics, ‘it’ refers to either a moment, a feeling, or an event that both the artist and the audience can relate to; again, this fosters connection and involvement (e.g. *‘It’s crazy’*, *‘It’s so true’*, *‘It’s time for me to have to go away’*, *‘It feels right’*). The frequent use of the pronoun ‘it’ is also a sign of a relatively inexplicit lexical content due to non-informational focus. Demonstrative pronouns, on the other

hand, are used to point to something specific, creating a sense of immediacy. By directing the audience's attention, the use of demonstrative pronouns enhances the connection and engagement with the content (e.g. *'That's why I'm by your side'*, *'This was gonna be our time'*, *'That's what I mean'*). On the other hand, the frequent use of indefinite pronouns in WBB&WGG pop song lyrics, "which substitute for fuller noun phrases" (Biber, 1988, p. 106), creates a universal appeal; as they make the lyrics relatable to a broad audience, as if the artist is speaking to or about everyone. Moreover, since indefinite pronouns do not refer to specified nouns, they do not carry much information (e.g. *'Everyone has their crazy world'*, *'Everybody knows how the story goes'*, *'Everybody needs a true romance'*, *'What's everybody talking about'*, *'Everybody needs some affection'*, *'Everybody sing'*, *'Nothing is like it used to be'*).

Similarly, the frequent use of emphatics (Z-Score = 1.48) in WBB&WGG pop song lyrics contributes to an involved, interactive discourse, enhancing the expressiveness and impact of the communication. Emphatics are linguistic devices that emphasize or intensify certain aspects of the message, drawing attention to specific words or ideas (Biber, 1988). In WBB&WGGPSL, they play a crucial role in creating emphasis, conveying emotions, and establishing a strong connection with the audience (e.g. *so, just, really, too*).

Additionally, the use of WH-questions (Z-Score = 0.78) marks involvement between the interlocutors. Generally, questions are considered as devices showing interest and involvement in discourse. According to Biber (1988), WH-questions are markers which indicate a high degree of interpersonal interaction and personal involvement; they "are used primarily in interactive discourse where there is a specific addressee present to answer questions" (p. 106). This contributes to a conversational tone in WBB&WGG pop song lyrics; as the use of WH-questions simulates a dialogue

between the artist and the listener, making the audience feel directly addressed. This conversational tone enhances the relatability of the lyrics, as it mirrors the way people communicate in everyday conversations (e.g. ‘*What’s the situation?*’, ‘*What’s going on?*’, ‘*What’s it gonna be?*’, ‘*Where did you go?*’).

Three subordination features are frequently used in WBB&WGG pop song lyrics: causative subordination (Z-Score = 1.8) (*because*, ‘*cause*, *cuz*), WH-clauses (Z-Score = 1.97), and conditional subordination (*if*) (Z-Score = 1.02). According to Biber (1988), causative and conditional subordination are considered as “markers of affect or stance”; as causative subordination is “[a] justification for actions or beliefs” and conditional subordination is “conditions for actions or beliefs” (p. 107). In WBB&WGG pop song lyrics, the use of causative subordination provides a rationale or reasoning behind the singer’s actions, beliefs, or emotions. It serves as a means of expressing the singer’s stance or perspective on a given situation, adding depth and authenticity to the lyrics (e.g., ‘*Now, I’m crying ‘cause I know you want me*’, ‘*Don’t wanna be a fool in your eyes ‘cause what we had was built on lies*’, ‘*I’m never gonna treat you bad ‘cause I never wanna see you sad*’). The use of conditional subordination in WBB&WGG pop song lyrics introduces a sense of contingency and potential outcomes. It reflects the singer’s stance on certain conditions or circumstances that may influence their actions or beliefs (e.g., ‘*I would treat you the best, if I knew then what I know now*’, ‘*How will I know, if I let you go*’). WH-clauses “serve interpersonal functions” (Biber, 1988, p. 232); they are often employed in WBB&WGG pop song lyrics to provide information. In interpersonal communication, this function is crucial for effective interaction. Providing details using WH-clauses promotes a communicative environment where participants share

knowledge (*'I can't believe what I just gave away now'*, *'You gotta believe what I say'*, *'I see how it's going down'*).

Discourse particles are also frequently used in WBB&WGG pop song lyrics (Z-Score = 0.15) (e.g. *well, alright, by the way, so, anyway*). Discourse Particles (or discourse markers) are lexemes which are syntactically independent and do not change the meaning of the sentence. They are loosely attached to the clause in order to facilitate the ongoing interaction. They serve to indicate the speaker's attitude and to make discourse more coherent. Carter and McCarthy (2006) argue that discourse markers "link segments of the discourse to one another in ways which reflect choices of monitoring, organization and management" (p. 208). Similarly, Biber et al. (1999, p. 1088) aver that discourse markers can act as monitoring devices, whereby the person who holds the conversational floor can check that other participants are still tuned into what is being said. Biber et al. (1999) argue that:

"Discourse markers are inserts which tend to occur at the beginning of a turn or utterance, and to combine two roles: (a) to signal a transition in the evolving progress of the conversation, and (b) to signal an interactive relationship between speaker, hearer, and message." (p. 1086).

The use of discourse particles in WBB&WGG pop song lyrics serves to provide insight into the speaker's attitude and contribute to the coherence of the lyrics. Moreover, just as in spoken discourse, WBB&WGG pop song lyrics benefit from markers that signal transitions between different sections or themes. These transitions mark shifts in the story being told, changes in mood, or variations in the overall message of the song. Additionally, discourse markers act as monitoring devices to ensure that the audience remains engaged. They serve as cues to capture attention or to emphasize key points in the lyrics, keeping the listeners connected with the song.

Hedges (Z-Score = 0.35) and possibility modals (Z-Score = 1.9) are also frequently employed in WBB&WGG pop song

lyrics. They are usually used to mark “uncertainty or lack of precision in the presentation of information” (Biber, 1988, p. 106). Hedges, which are a subclass of adverbs, are mitigating devices which mark propositions as probable or uncertain; as they indicate imprecision (Lakoff, 1973). Chafe (1982) argues that they convey doubt and mark fuzziness in involved discourse. Different hedges are frequently used in WBB&WGG pop song lyrics (e.g. *maybe, perhaps, kinda, sorta*). Similarly, possibility modals (*can, could, may, might*) are frequently used in WBB&WGG pop song lyrics and are markers of uncertainty. They are also a characteristic of involved discourse; they contribute to the emotional depth and the richness of the lyrical content, as they communicate a wide range of possibilities, allowing artists to evoke complex feelings and engage listeners on a deeper level.

Analytic Negation (Z-Score = 2.05), *be* as main verb (Z-Score = 0.2), and non-phrasal coordination (Z-Score = - 0.34) are frequently employed in WBB&WGG pop song lyrics, except for non-phrasal coordination which is less frequently used in the lyrics. These features are associated with “a fragmented presentation of information, resulting in a low informational density” (Biber, 1988, p. 106). Analytic negation is “an alternative to the more integrative synthetic negation (*no, either*)” (Biber, 1988, p. 106) (e.g. ‘*It isn’t there anymore*’, ‘*I don’t deserve to suffer*’, ‘*I don’t wanna be a fool*’, ‘*I just can’t fight*’). According to Biber (1988), *be* as main verb is frequently found in less complex constructions, indicating a reduced informational load; it “is used to modify a noun with a predicative expression, instead of integrating the information into the noun phrase itself” (p. 106) (e.g. ‘*You’re the first and last thing on my mind*’, ‘*This love is unbreakable*’, ‘*Together we’re strong*’, ‘*You’re beautiful*’). Non-phrasal *and* “is used to string clauses together in a loose, logically unspecified manner, instead of integrating the information into fewer units through the use of prepositional phrases, relative clauses, adjectives,

etc” (Biber, 1988, p. 106) (e.g. ‘*The sun goes down, and the light comes up*’, ‘*I need to turn around, and I want to recognize myself*’, ‘*I was your girl, and I fell for you*’).

Three additional linguistic features were identified in the WBB&WGGPSL corpus. These features —namely, the use of interjections, the use of base form verbs (most of which are used in imperative sentences), and first-person subject omission— were not originally accounted for in Biber (1988), rather the researcher identified their frequent occurrence throughout the corpus of the present study. Two of these features were already counted by MAT as tags of the Stanford Tagger since they were not included in the linguistic features specified in Biber (1988); these are interjections (Z-Score = 2.33) and base form verbs (Z-Score = 2.02). The third linguistic feature, first-person subject omission, was not counted by MAT, rather it was observed by the researcher during the editing process of MAT’s tagged texts. The frequent appearance of these three linguistic features significantly contributes to the involved aspect of Dimension 1. Interjections, as spontaneous and emotionally charged expressions, play a fundamental role in creating an involved discourse; the frequent use of interjections in WBB&WGG pop song lyrics introduces a heightened sense of immediacy and emotional connection between the artist and the audience (e.g., *Yeah, Oh, Aah, Whoa, Woo*). Incorporating interjections among the positive linguistic features of Dimension 1 acknowledges the role of emotionally charged expressions in enhancing the involvement aspect in WBB&WGG pop song lyrics through enriching the engagement between the artist and the listener.

The use of base form verbs, especially in imperative sentences in WBB&WGG pop song lyrics, whether addressing the listener directly or another character or entity within the lyrics, acts as an indication of involvement and contributes to the overall emotive and engaging quality of the lyrics. It reflects the songwriter’s intent to create a sense of connection

and direct engagement with the listener. Imperative sentences typically convey commands, requests, or suggestions. Songwriters express concise and straightforward directives through employing base form verbs. This makes the listener feel directly addressed and involved in the song. For example, clauses such as ‘*Sing it*’ and ‘*Dance with me*’ utilize base form verbs in order to convey a sense of urgency and participation which creates an interactive experience for the audience. Similarly, commands addressed to a character within the song lyrics, such as ‘*Hold me tight,*’ ‘*Stay a while,*’ ‘*Let it go,*’ ‘*Show me the meaning of being lonely,*’ ‘*Tell me why I can’t be there where you are,*’ ‘*Keep shining brightly,*’ ‘*Let me take you there,*’ and ‘*Show me all your love and walk right through*’ convey a sense of personal connection and involvement with the addressed person, which enriches the emotional depth and involvement of the song.

First-person subject omission (*I*), a linguistic feature often prevalent in song lyrics, contributes significantly to the involved pole of Dimension 1. “Omission is a term in grammatical analysis which means an omission of a part of structure in a sentence” (Crystal, 2008, p. 166). It refers to the deliberate exclusion of a part of the sentence structure, which means that a sentence with omission is not a complete sentence (e.g. ‘*Guess that I was blind*’, ‘*Wish I could take back letting you go*’, ‘*Can hardly believe it*’). In standard grammar, a complete sentence typically includes both a subject and a verb, yet in such constructions, the subject is implied rather than explicitly stated. As a unique form of linguistic expression, WBB&WGG pop song lyrics often deviate from conventional grammatical structures in order to convey emotions and share experiences. Unlike traditional written or spoken language, songwriters frequently employ first-person subject omission to enhance brevity, rhythmic flow, intimacy, and simplicity. Additionally, subject omission aligns with the informality of

colloquial language, which contributes to the authenticity of the lyrics.

Through recognizing the use of interjections, base form verbs (in imperative sentences), and first-person subject omission in WBB&WGG pop song lyrics as three linguistic features apparent in the present corpus, the study extends Biber's (1988) linguistic framework, specifically Dimension 1 - Involved vs. Information Production, enhancing the involved aspect of Dimension 1. Interjections show spontaneity and emotional immediacy, first-person subject omission enhances brevity and rhythmic cohesion, and the use of base form verbs in imperative sentences signifies involvement and contributes to the overall emotive impact and involvement of the lyrics.

On the other side, the lyrics in the present corpus are characterized by the lack of most of the features that belong to the negative pole of this dimension. The two most prominent features characterizing the lyrics as having an involved style as opposed to a more informational style are low type/token ratio (TTR) (Z-Score = - 5.3) and low average word length (AWL) (Z-Score = - 2.36). These two features determine the lexical density of texts. Biber et al. (1999, p. 62) state that lexical density refers to the proportion of the text made up of lexical word tokens (nouns, lexical verbs, adjectives, and adverbs). TTR is a measure that assesses the lexical richness or diversity of a text. The low TTR suggests that the corpus has a high level of repetition and relies on a limited set of vocabulary. Repetition is a common stylistic feature in songwriting, where certain phrases or lines are intentionally repeated. Songs contain repeated sections like choruses and bridges. This repetition contributes to a lower TTR but serves specific artistic purposes, such as emphasizing key themes, enhancing rhythm, creating an emotional impact, or reinforcing memorability. This creates a connection with the audience; listeners can easily participate by singing along, as the repetitive nature contributes to the song's appeal.

On the other hand, Biber (1988) argues that the use of longer words in texts conveys specific and specialized information. The low AWL in the present corpus indicates that the words used in the lyrics are relatively short. This could be attributed to the fact that one of the goals of the songwriter is often to create lyrics that are not overly complex linguistically but are still rich in emotional expression. Moreover, shorter words are often more accessible and easier to understand and memorize for a wider audience; as they can convey messages more directly and clearly.

Exploring how this level of involvement compares to other genres within Biber's (1988) framework provides insights into the relation among WBB&WGGPSL and other genres with respect to Dimension 1: Involved versus Informational Production. The WBB&WGG pop song lyrics corpus has a high level of involvement which is different from all genres apart from Conversations: Telephone Conversations (D-Score = 37.2) and Face-to-face Conversations (D-Score = 35.3), which are identified as the closest genres to the present corpus in terms of involvement. These genres have notably positive scores on Dimension 1, which suggests that they align more closely with the interactive, affective, and involved end of this dimension. Other genres like Personal Letters (D-Score = 19.5), Spontaneous Speeches (D-Score = 18.2), Interviews (D-Score = 17.1), Prepared Speeches (D-Score = 18.2), and Romantic Fiction (D-Score = 4.3) have less positive scores on this dimension. On the other hand, the WBB&WGG pop song lyrics corpus differs significantly from several other genres in Biber's (1988) framework. These genres have negative scores on Dimension 1, indicating a high informational focus and a careful integration of information in the texts. These genres are Official Documents (D-Score = -18.1), Academic Prose (D-Score = -14.9), Press Reportage (D-Score = -15.1), etc.

4.2 Dimension 2: Narrative versus Non-Narrative Concerns

Dimension 2: Narrative versus Non-narrative Concerns distinguishes narrative discourse from other types of discourse. It includes ten linguistic features, six positive and four negative features. The positive markers, which are considered as markers of narrative action, are past tense verbs, third person pronouns, perfect aspect verbs, etc. The negative markers for narratives are present tense verbs, attributive adjectives, past participial whiz deletions, and word length. High scores on this dimension represent active, event-oriented, narrative discourse as opposed to more descriptive or expository, non-narrative types of discourse which are associated with low scores on this dimension. The linguistic features which belong to the narrative end of this dimension - past tense verbs, third person personal pronouns, perfect aspect verbs, public verbs, etc. are markers of narrative action and contribute to a dynamic storytelling style. On the contrary, negative markers for narratives include present tense verbs and attributive adjectives, among others. These features are associated with non-narrative concerns, reflecting a more descriptive form of discourse.

The WBB&WGGPSL Corpus has a Dimension 2 score of - 2.34 (Figure 2) which is a moderately low score. This score suggests that most of the song lyrics in this corpus lack many of the positive features associated with Dimension 2, indicating a departure from a narrative-oriented discourse style.

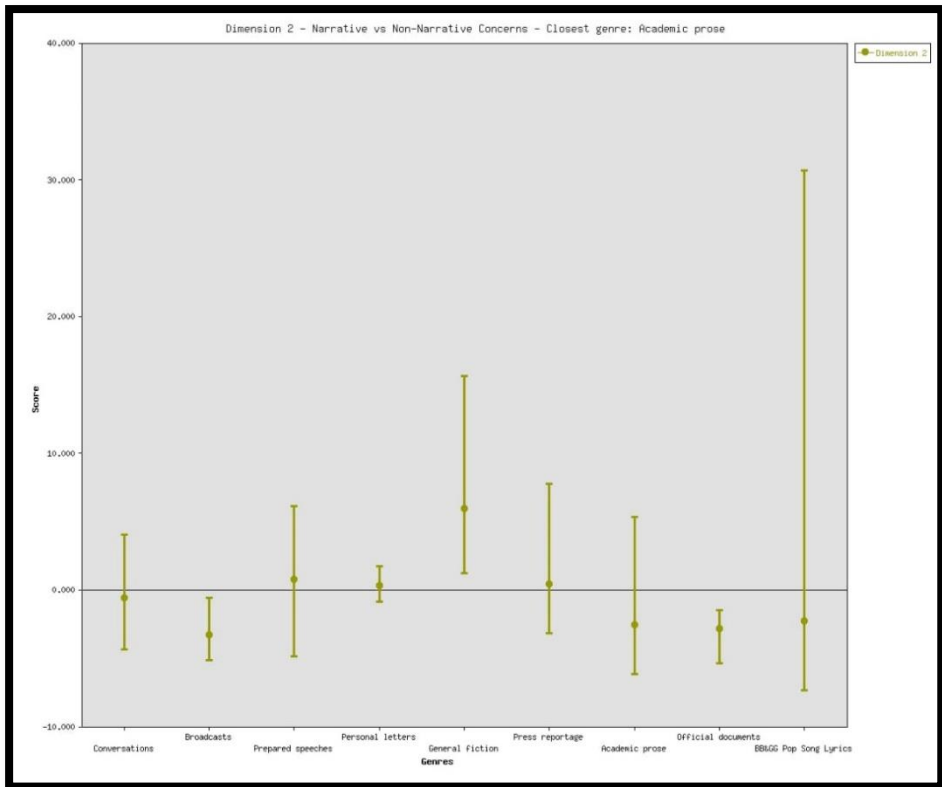


Figure 2. MAT's visualization of the WBB&WGGPSL Corpus along Dimension 2

The absence of these narrative markers suggests that WBB&WGG pop song lyrics are not characterized by a strong narrative focus. Conversely, the WBB&WGGPSL Corpus exhibits relatively high scores for some features associated with the negative pole of Dimension 2 (non-narrative concerns). Notably, present tense verbs are prominent in the corpus, which aligns with the unmarked value of non-narrative concerns for this dimension. Attributive adjectives, on the other hand, occasionally appear in the corpus, but less frequently.

The use of present tense verbs (Z-Score = 0.94) in WBB&WGG pop song lyrics serves a specific communicative purpose related to conveying emotions and details about relationships. This suggests a real-time or immediate engagement with the audience, as if the artist is sharing details

about an ongoing relationship or experience. Conversely, song lyrics with low frequencies of present tense verbs and relatively high scores of past tense verbs usually tend to focus on recounting past events, which is a characteristic of a storytelling approach where events are being revisited rather than actively unfolding in the current moment. Thus, present tense verbs create immediacy, conveying a sense of timelessness and universal relevance. By describing situations in the present, WBB&WGG pop song lyrics have an impact on listeners across different times and cultures, making the themes expressed in the lyrics more empathetic and permanent. Moreover, the continuous nature of the present tense contributes to a feeling of ongoing and evolving experiences. Past tense verbs, on the other hand, evoke a sense of retrospection, which contributes to a storytelling style that reflects on shared experiences or memories.

Attributive adjectives are less frequently used in WBB&WGG song lyrics (Z-Score = - 1.87). They are markers of non-narrative concerns, as they contribute to a descriptive and expressive style rather than a narrative-driven structure. Attributive adjectives are used to provide additional description and details about nouns in song lyrics. When these adjectives are prevalent, they often indicate a focus on creating vivid images rather than advancing a narrative. Moreover, the use of attributive adjectives introduces a level of subjectivity into the lyrics. Instead of objectively narrating events, the emphasis is on the artist's interpretation, perception, and emotional response to the subject matter in the song (e.g. '*Our happy home*', '*Our own funny moods*', '*My foolish pride*', '*A green hill*', '*That old oak tree*', '*Your broken heart*', '*The hardest day*', '*My crazy mind*').

Several genres in Biber's (1988) framework have negative scores on Dimension 2 similar to WBB&WGG pop song lyrics. These include Academic Prose (D-Score = - 2.6), Broadcasts (Z-Score = - 3.3), Hobbies (D-Score = - 2.9),

Official Documents (D-Score = - 2.9), Professional Letters (D-Score = - 2.2), Telephone Conversations (D-Score = - 2.1), Interviews (D-Score = - 1.1), and Face-to-face Conversations (D-Score = - 0.6). Academic Prose has the closest Dimension 2 score to WBB&WGG pop song lyrics, suggesting some similarity in the non-narrative aspects of discourse. This comparison highlights that WBB&WGGP song lyrics share certain linguistic features with academic writing, possibly in terms of descriptive or non-narrative elements. The rest of Biber's (1988) genres, which include: Romantic Fiction (D-Score = 7.2), Mystery Fiction (D-Score = 6.0), Science Fiction (D-Score = 5.9), Adventure Fiction (D-Score = 5.5), etc., have positive scores on *Dimension 2*. This indicates that these genres align more closely with narrative concerns and are different from the style observed in the WBB&WGGPSL Corpus.

4.3 Dimension 3: Explicit Versus Situation-Dependent Reference

Dimension 3: explicit versus situation-dependent reference distinguishes between highly explicit, context-independent reference and nonspecific, situation-dependent reference. It contains eight linguistic features, including three relative clauses, *wh* relative clauses on object positions, pied piping relative clauses, and *wh* relative clauses on subject positions. Negative markers for explicit context-independent texts are time adverbials, place adverbials and adverbs. These features characterize context-dependent references, as they depend on the context for interpretation and are less explicit.

The WBB&WGGPSL Corpus has a Dimension 3 score of -5.18 (Figure 3), which is a low score indicating a notable tendency towards employing linguistic features associated with situation-dependent references in contrast to explicit references. This low score is attributed not only to the frequent occurrence of markers of discourse-dependent reference but

also to the infrequency of markers of explicit reference in WBB&WGG pop song lyrics.

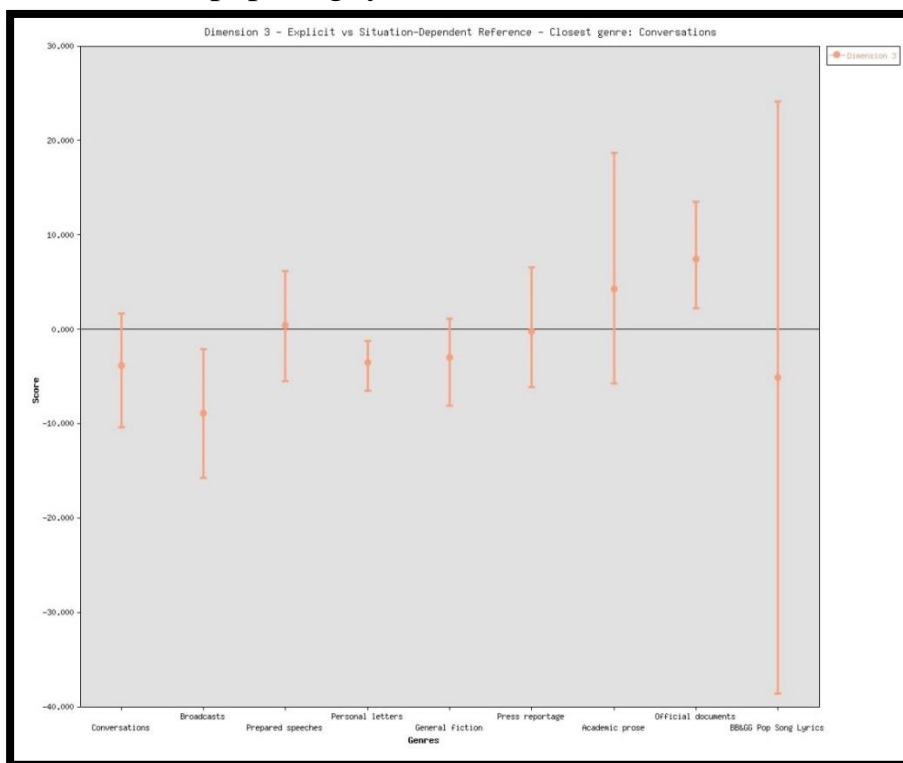


Figure 3. MAT's visualization of the WBB&WGGPSL Corpus along Dimension 3

Time and place adverbials (Z-Scores = 1.73 and 1.19, respectively) are two of the linguistic features that mark situation-dependent reference in the corpus under investigation (e.g. *then, today, yesterday, now, tomorrow, tonight, above, away, far, here, near*). Biber (1988) argues that time and place adverbials can act as deictic expressions, requiring knowledge of the physical or temporal context in order to be interpreted (p. 110). They are used for locative and temporal reference. Although they can be used for text-internal referents, they are more commonly used for reference to places and times outside of the text itself. These adverbials contribute to creating vivid

imagery, setting, and temporal context in the song lyrics. For instance, time adverbials like ‘yesterday’ and ‘tonight’ evoke specific moments in time, while place adverbials like ‘here’ and ‘far’ establish a sense of place. The deictic function of these adverbials adds layers of meaning and emotions, as listeners connect with the specific times and places evoked by the lyrics. Moreover, Chafe and Danielewicz (1987) consider place and time adverbials as markers of involvement in discourse. On the other hand, adverbs are less frequently used in the present corpus (Z-Score = -0.74) (e.g. *too, rather, again, never*). Adverbs in the corpus modify verbs, adjectives, or other adverbs, which influences the tone and intensity of the expressions in the lyrics.

The closest genre to the present corpus in Biber’s (1988) analysis is Telephone Conversation (D-score = - 5.2) which shares the same tendency towards context-dependent reference. Other genres that have negative scores on this dimension are Broadcasts (D-Score = - 9.0), Romantic Fiction (D-Score = - 4.1), Face-to-face Conversations (D-Score = - 3.9), Adventure Fiction (D-Score = - 3.8), Mystery Fiction (D-Score = - 3.6), Personal Letters (D-Score = - 3.6), General Fiction (D-Score = - 3.1), Science Fiction (D-Score = - 1.4), Humor (D-Score = - 0.8), Interviews (D-Score = - 0.4), and Press Reportage (D-Score = - 0.3). These genres similarly exhibit a reliance on context for interpretation, emphasizing implicit reference. On the contrary, genres with positive scores on this dimension, including Official documents (D-Score = 7.3), Professional Letters (D-Score = 6.5), Press Reviews (D-Score = 4.3), Religion (D-Score = 3.7), Press Editorials (D-Score = 1.9), Spontaneous Speeches (D-Score = 1.2), and others, suggest a preference for explicit references, contributing to clearer and more direct communication.

5. Conclusion

This study conducted a multi-dimensional analysis of WBB&WGG song lyrics. The corpus of the present study was analyzed along the first three dimensions of Biber's (1988) six dimensions of variation. Dimension 1 represents a continuum between affective, interactional, and generalized content on one end, and high informational density and exact informational content on the other. According to Biber (1988), this dimension is influenced by two key communicative parameters: the primary purpose of the writer/speaker (informational versus interactive, affective, and involved) and the production circumstances (careful editing with precision in lexical choice and integrated textual structure versus real-time constraints leading to generalized lexical choice and fragmented information presentation). WBB&WGG pop song lyrics often emphasize conveying emotions, personal experiences, and establishing a connection with the audience. This aligns with the affective and involved aspect of communication described in this dimension. The lyrics usually focus on expressing feelings or describing an experience rather than conveying specific, detailed information. While the process of songwriting does not often occur under immediate pressure of real-time constraints, it is characterized by a spontaneous and conversational structure. This often leads to a generally fragmented presentation of information within the lyrics. The emphasis in songwriting is occasionally placed on capturing the essence of a particular moment or emotion, rather than thoroughly crafting a well-structured and formalized text. The result is a style that emphasizes authenticity and emotional involvement, allowing for a more immediate connection between the artist and the audience. This is achieved through the use of linguistic features that contribute to a distinctive style marked by involvement, brevity, and spontaneity.

The results of the analysis of Dimension 2 in the WBB&WGGPSL Corpus suggest that most of the song lyrics

tend to veer away from a narrative discourse style. The absence of key narrative markers and the presence of features associated with non-narrative concerns contribute to a discourse that is less event-oriented and more aligned with descriptive forms of expression.

Finally, the results of the analysis of Dimension 3 in the WBB&WGGPSL Corpus suggest that most of the song lyrics tend to deviate from explicit, situation-independent references. The absence of key explicitness and context-independent markers and the presence of features associated with situation-dependent references contribute to a discourse that inclines towards the use of indirect, context-dependent expressions, opting for a style that invites interpretation and relies on the surrounding context for a fuller understanding. This tendency towards veering away from explicitness aligns with the artistic and emotive nature of WBB&WGG pop song lyrics, intensifying a richness of meaning that invites listeners to engage with and interpret the lyrics in a more subjective and personal manner.

References

- Bertoli-Dutra, P. (2014). Multi-Dimensional analysis of pop songs. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis, 25 Years on: A Tribute to Douglas Biber* (pp. 149–175). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.60>
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Bridle, M. (2011). *Male Blues Lyrics 1920-1965: A Corpus Assisted Analysis* [Master's thesis, University of Huddersfield]. Eprints. <http://eprints.hud.ac.uk/id/eprint/10900/>
- Carter, R, & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide: Spoken and Written English Grammar and Usage*. Cambridge University Press. https://rafieienglishclinic.com/upload/blog/Ronald_Carter_Michael_McCarthy_Cambridge%20Grammar%20of%20English_Cambridge%20University%20Press%20%282006%29.pdf
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In Tannen, D (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35-54). Ablex.
- Chafe W., & J. Danielewicz (1987). Properties of Spoken and Written Language. In R. Horowitz, & J. Samuels (Eds.), *Comprehending Oral and Written Language* (pp. 83-113). Academic Press. <https://files.eric.ed.gov/fulltext/ED282230.pdf>

- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Blackwell Publishing. <http://dx.doi.org/10.1002/9781444302776>
- Falk, J (2012). *We Will Rock You: A Diachronic Corpus-based Analysis of Linguistic Features in Rock Lyrics*. [Bachelor's Thesis, Linnaeus University]. Diva Portal. <https://www.diva-portal.org/smash/get/diva2:605003/FULLTEXT02.pdf>
- Kreyer, R., & Mukherjee, J. (2007). The Style of Pop Song Lyrics: A corpus-linguistic pilot study. *Anglia: Journal of English Philology*, 125(1), 31-58. <https://doi.org/10.1515/angl.2007.31>
- Lakoff, G. (1973). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4), 458-508.
- Ma, S. (2017). *A multidimensional analysis of lyrics of modern English popular songs*. [Master's thesis, University of Hong Kong]. The HKU Scholars Hub. <https://hub.hku.hk/handle/10722/252027>
- Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 67-94). Bloomsbury Academic Publishing PLC. <https://dx.doi.org/10.5040/9781350023857.0012>