



P-ISSN 2355-2794
E-ISSN 2461-0275

Exploring Vocabulary Characteristics across Nine Music Genres: A Corpus-Based Study for Vocabulary Learning

Nonthawat Nonthasen
Sinaphat Samakketwit
Patsarapa Srisompak
Piyapong Laosrirattanachai*

English for Service Industry Programme, Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, THAILAND

Abstract

The use of music for vocabulary learning has been extensively researched, yet a gap remains in studies focusing on vocabulary across different music genres. This quantitative corpus-based study aims to examine the vocabulary characteristics of nine music genres and their implications for vocabulary learning. To explore these characteristics, lexical profiling, lexical level, CEFR mapping, lexical variation, and lexical density of songs from each genre were examined. Data were collected from Billboard and YouTube songs categorised into nine genres from 2014 to 2023, forming a corpus of 671,051 tokens. The analysis was conducted using AntWordProfiler and VocabProfile. The results revealed that jazz had the highest GSL coverage (91.33%), while hip-hop recorded the lowest (78.88%) but had the highest OWL usage (20.52%). Country and folk (95.20%), jazz (94.74%), and alternative (94.23%) exhibited the greatest use of high-frequency vocabulary, whereas hip-hop displayed the lowest (87.01%) alongside the highest mid- (4.06%) and low-frequency proportions (8.93%). CEFR analysis indicated that country and folk, jazz, alternative, and rock contained mostly A1–A2 vocabulary; R&B/soul, children's, pop, and electronic genres reflected B1–B2 levels; and hip-hop showed the highest C1–C2 coverage. Jazz (43.44%) and children's songs (39.73%) demonstrated the greatest lexical variation, while hip-hop (52.88%), children's (51.80%), and R&B/soul (50.25%) exhibited the highest lexical density. These findings provide guidance for selecting music genres to enhance vocabulary learning at different proficiency levels both inside and outside the classroom.

* Corresponding author, email: piyapong.l@ku.th

Citation in APA style: Nonthasen, N., Samakketwit, S., Srisompak, P., & Laosrirattanachai, P. (2025). Exploring vocabulary characteristics across nine music genres: A corpus-based study for vocabulary learning. *Studies in English Language and Education*, 12(3), 1509-1526.

Received September 22, 2024; Revised January 27, 2025; Accepted September 15, 2025; Published Online September 30, 2025

<https://doi.org/10.24815/siele.v12i3.41375>

Copyright © 2025 by Authors, published by *Studies in English Language and Education*. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>)

Keywords: Education, music genre, proficiency level, vocabulary characteristics, vocabulary learning.

1. INTRODUCTION

English language learning today encompasses a multitude of methods, both inside and outside the classroom. Whether learning takes place through textbooks or from the surrounding environment, acquiring vocabulary remains central, as it underpins the fundamental language skills of listening, speaking, reading, and writing (Coxhead, 2000). Remembering and learning a large number of vocabulary words can, in fact, be challenging due to the continuous emergence of new words and the sheer volume of existing vocabulary. This vocabulary acquisition can be approached in various ways, including the use of music as a tool for learning English vocabulary in the classroom. For instance, a study by Tomczak and Lew (2019) recommended the early introduction of music in language classrooms, particularly as song lyrics aided in vocabulary learning. Moreover, the use of music in teaching English as a Second Language (ESL) has also been proven to be effective (Tasnim, 2022; Zhang et al., 2023).

In addition, learning in a stress-free environment, such as through music, can enhance learner engagement and positively affect learning outcomes (Krashen, 1982). Music, which is rich in diverse vocabulary, helps learners acquire a substantial vocabulary for effective communication in a second language and exposes them to new words they might not encounter in the classroom or textbooks. The use of single words and repeated phrases in many songs serves as effective input for second language (L2) learning (Webb, 2019), and the melody, catchy phrases, rhythm, and beats help listeners remember vocabulary more effectively (Sadiqzade, 2025). One of the fundamental aspects that requires consideration in research on songs is the categorisation of musical genres. For instance, Schultz and Huron (2015) identified 12 genres, including avant-garde, blues, classical, country, folk, jazz, pop/rock, rap, R&B, reggae, religious, and theatre. Schreiber (2015) proposed a song classification of 10 genres, comprising rock, pop, alternative, jazz, hip-hop, hip-hop/rap, soundtrack, R&B, electronic, and country. Deja et al. (2016) also identified 10 genres, such as gospel, rock, soft rock, acoustic, rap, children's, Christmas, jazz, reggae, and special. These approaches to genre classification in the aforementioned studies were all based on rhythmic structure, harmonic content, and instrumentation (Tzanetakis & Cook, 2002).

Although the use of music for vocabulary learning has been widely studied, no specific research has examined vocabulary acquisition related to different music genres for learners at varying proficiency levels. A review of the literature revealed that specific song genres have been studied for their positive impact on language learning. The examples include the folk genre (Spicher & Sweeney, 2007), the pop genre (Tegge, 2018), and the hip-hop genre (Kao & Oxford, 2014). However, no study has comprehensively explored all genres of music since different genres of music may be more suitable for learners at different proficiency levels. This study aims to identify the most effective genres of music for learners with varying levels of language proficiency by analysing the characteristics of vocabulary to evaluate the complexity of song lyrics. Vocabulary characteristics can be investigated in various ways. A review of the literature indicated that common approaches include lexical profiling (Coxhead & Hirsch, 2007; Laosrirattanachai & Laosrirattanachai, 2025), lexical level (Laosrirattanachai & Laosrirattanachai, 2023; Nation, 2022), CEFR (Common European Framework of Reference for Languages) (Loesnophaimongkhon et al., 2025), lexical variation (Laosrirattanachai et al., 2025; Treffers-Daller et al., 2018), and lexical density (Liu, 2021; Meebangsai et al., 2023). Each of these methods highlights different aspects of vocabulary within a corpus.

Lexical profiling has the concept that each word should belong to a single profile. This approach analyses vocabulary groups by categorising words into relevant clusters and excluding those that are irrelevant (Laosrirattanachai & Laosrirattanachai, 2021; Na Ayutthaya et al., 2022). Initially, lexical profiling relied on two reference lists (Laufer & Nation, 1995); the General Service List (GSL), containing 2,000 high-frequency words and covering about 80% of texts

(Nation & Waring, 1997; West, 1953); and the Academic Word List (AWL), comprising 570 word families relevant to academic contexts and accounting for around 10% of texts (Coxhead, 2000). Words not included in these lists were classified under the Outside Word List (OWL), consisting of low-frequency and specialised terms, also representing about 10% of texts (Coxhead & Hirsch, 2007).

Lexical level refers to the 26 stages of vocabulary determined by word frequency (Nation, 2016). These levels ranged from 1,000 to 25,000 words, denoted as K1-K25, where K1 comprises the most frequently used words, and K25 contains the least common or more specialised terms. Vocabulary beyond K25, which typically consists of very low-frequency words, is classified as the off-list level. Schmitt and Schmitt (2014) divided the K1-K25 into three segments, including high-frequency words (commonly used in everyday language) from K1 to K3 (1 to 3,000), medium-frequency words (less common, but used in specific contexts) from K4 to K9 (3,001 to 9,000), and low-frequency words (rarely used and often technical) from K10 to K25 (9,001 to 25,000). Lexical level is also linked to lexical coverage, indicating the extent to which a reader comprehends a text based on their vocabulary knowledge (Schmitt et al., 2011). Lexical coverage is usually expressed as a percentage of a corpus accounted for by known word lists. Achieving 95% comprehension allows for grasping main ideas, whereas 98% comprehension ensures a deeper understanding of details and nuances (Schmitt et al., 2011).

The Common European Framework of Reference for Languages (CEFR), developed by the Council of Europe, is a significant framework for language learning. It served multiple purposes, including the development of examinations and language courses (Council of Europe, 2001). The CEFR comprises five core language skills: listening, reading, spoken interaction, spoken production, and writing (Krishnan & Yunus, 2019). It is categorised into six levels of language proficiency, from beginner to advanced. Levels A1 and A2 denote a 'Basic User', B1 and B2 indicate an 'Independent User', while C1 and C2 represent a 'Proficient User' (Ito, 2019). In studies on vocabulary levels in relation to the CEFR, Nation and Crabbe (1991) proposed a framework for mapping lexical levels to CEFR levels as follows: CEFR A1 corresponded to a vocabulary of around 120 words, while A2 expanded that to about 2,000 words. At the B1 level, a learner would know around 3,000 words, and at the B2 level, roughly 4,000 words. C1 proficiency involved knowing about 5,000 words, and C2 indicated a vocabulary of 6,000 words or more.

Lexical variation refers to the range of vocabulary used in a text, highlighting the importance of diverse word use to enhance richness (Schmitt & Schmitt, 2020). Learners' vocabulary breadth and diversity reflect their proficiency levels (Jarvis, 2013) and contribute to text quality, as greater lexical variation indicates sophistication, whereas frequent repetition signals limited vocabulary (Johansson, 2008). To quantify lexical variation, Laufer and Nation (1995) proposed the Type-Token Ratio (TTR), in which a 'type' is each unique word in a text, while a 'token' is any individual word occurrence. However, TTR has limitations, particularly for longer texts, as it tends to decline due to increased word repetition (Gayraud, 2000; Kyle et al., 2021). Despite this, TTR remains a valuable measure for analysing lexical variation in songs, which are generally short and comparable in length, mitigating the issue of TTR decline. Therefore, TTR provides an effective means of assessing vocabulary diversity in musical texts while accounting for text length constraints.

Finally, lexical density measures the proportion of content words relative to the total number of words in a text (Thornbury & Slade, 2006). A high proportion of content words typically indicates detailed and rich content. Most spoken texts exhibited a lexical density below 40%, whereas written texts generally reached a lexical density of 40% or higher (Ure, 1971). This distinction illustrates that written texts are more informative than spoken texts, as spoken language frequently employs more connectors. Texts that contain a higher concentration of content words provide more opportunities for encountering new vocabulary.

Based on the Input Hypothesis theory (Krashen, 1982), language learning is most effective when learners are exposed to comprehensible input that is slightly beyond their current level of knowledge ($i+1$) yet still understandable, thereby facilitating progress to the next level, known as $i+1$. In light of this principle, when selecting songs from different genres for language learning,

learners should consider genres that correspond to an appropriate level of difficulty. The five aspects used to analyse vocabulary characteristics can reveal the lexical features of each genre, and thus support more informed and suitable choices. Based on the identified research gap, this study aims to address the following research questions:

1. What percentage of lexical profiling, including the GSL, AWL, and OWL, is represented in nine corpora of English songs?
2. How are high-, mid-, and low-frequency words distributed across the nine corpora of English songs in relation to lexical levels (K1-K25 and off-list words), and what does this reveal about the overall lexical coverage of the corpora?
3. At what CEFR level should songs in each genre be categorised?
4. What is the TTR of lexical variation observed in nine corpora of English songs?
5. What is the lexical density ratio in nine corpora of English songs?

2. METHODS

2.1 Research Design

This study employed a quantitative corpus-based approach. A corpus study provided authentic language data that can be systematically analysed using specialised software. Corpus linguistics was recognised as a reliable method in fields such as language learning, discourse analysis, and vocabulary research (O’Keeffe & McCarthy, 2010), offering evidence-based insights into naturally occurring language use (Crystal, 1991). A quantitative design was chosen to examine lexical characteristics, such as profile, level, CEFR, variation, and density, across nine music genres. Unlike qualitative methods, corpus-based quantitative analysis can process large datasets objectively and reveal statistically significant patterns.

2.2 Data Source

The data for this study were obtained primarily from songs listed in the Billboard Hot 100 chart, a widely recognised platform that tracks popular songs based on sales and airplay. The Billboard chart was selected because of its credibility, global influence, and comprehensive coverage of commercially successful music. Songs listed weekly in the Billboard Hot 100 chart from the first week of January 2014 to the last week of December 2023 were identified, recorded, and subsequently classified into their respective genres for analysis. These sources provided access to mainstream music consumption patterns over a sustained period, making them highly suitable for examining vocabulary use across diverse song genres. For music genres that did not meet the required number of songs listed in the Billboard Hot 100 chart, songs from YouTube within the same year range, with a minimum of 100,000 views, were included to compensate for the shortfall.

2.3 Data Collection

In this research, we categorised genres based on similarities in rhythm and melody, as well as differences in tempo (Foleiss & Tavares, 2019; Wundervald & Zeviani, 2019), and identified nine genres: hip-hop, pop, rock, R&B/soul, country and folk, children’s, alternative, electronic, and jazz. For each genre, 200 songs were collected. Priority was given to songs ranked in the Billboard Hot 100 from 2014 to 2023. However, certain genres had fewer than 200 songs listed in the Billboard Hot 100 due to their niche audiences. For example, only 10 jazz songs appeared on the chart, while the children’s genre contained 84 songs. Although relatively few jazz songs charted over the ten years, the genre’s historical significance confirmed its enduring popularity and dedicated followers (Early & Monson, 2019). Similarly, children’s songs could foster strong learner motivation, as young learners expressed particular enjoyment in learning and singing them (Sukirmiyadi, 2017). Accordingly, both jazz and children’s songs were included as primary

genres in this study. Jazz and children's songs were additionally collected from YouTube to address the shortfall. YouTube's algorithm typically recommended videos based on users' listening behaviour, selecting tracks that are similar to those previously played. To avoid selection bias, songs were chosen from those randomly recommended by the platform. Data collection continued until 200 songs were obtained for genres that could not be fully populated using the Billboard Hot 100 alone. A summary of the dataset is presented in Table 1.

Table 1. Details of the nine song corpora.

Genre	Number of songs	Number of tokens			
		Max	Min	Average	Total
Hip-hop	200	1,282	168	616	122,778
Pop	200	990	91	409	80,906
Country and folk	200	583	94	336	66,724
Rock	200	1,064	115	348	69,746
Electronic	200	961	98	410	81,846
R&B/soul	200	1,102	125	472	93,193
Alternative	200	809	98	319	63,509
Children's	200	1,037	37	277	55,593
Jazz	200	740	51	195	36,756
Total	1,800				671,051

2.4 Data Analysis

In addition to selecting appropriate data sources and collection methods, the choice of analytical tools was also crucial. In this study, corpus tools were employed for analysis, including AntWordProfiler (Anthony, 2024), a lexical analysis tool that profiles vocabulary in texts by comparing words against reference lists, such as the GSL and AWL. It also calculates types, tokens, and lexical coverage. Another tool, the VocabProfile (Cobb, 2022), a computer programme that analyses a text and categorises its words into frequency-based levels, each comprising 1,000 reference words, was utilised to analyse the lexical level. It showed the percentage of words that fall into each level, ranging from level 1 to level 25, and the off-list level.

Firstly, lexical profiling was conducted by categorising vocabulary into three profiles: the GSL, the AWL, and the OWL. The GSL and AWL served as reference lists, while words not found in these lists were classified under OWL. Analysing the distribution of vocabulary across these profiles provided insights into whether songs within each genre demonstrated similar or differing proportions of vocabulary usage. In this study, a lexical profiling approach was adopted, utilising the GSL, AWL, and OWL, to classify vocabulary within song lyrics. Genres with a high prevalence of GSL words were deemed suitable for learners with lower English proficiency, while those rich in AWL or OWL vocabulary were more appropriate for advanced learners (Laosrirattanachai & Ruangjaroon, 2020, 2021; Nation, 2016).

Next, the lexical level of vocabulary was analysed by classifying it into 26 levels, ranging from K1 to K25 and off-list. Lower-level vocabulary, such as K1, consisted of common words typically suitable for beginner level. Higher-level vocabulary was categorised progressively from K1 to K25 based on increasing complexity. Additionally, the vocabulary was grouped into three frequency categories: high-frequency words (K1-K3), mid-frequency words (K4-K9), and low-frequency words (K10-K25). In this study, genres featuring lower-level vocabulary were considered suitable for learners with less advanced English proficiency, whereas those with higher-level vocabulary were better suited for more proficient learners.

The study also incorporated a CEFR analysis, adapting the levels proposed by Nation and Crabbe (1991). To map the CEFR levels (A1-C2) with the lexical levels (K1-K25) used in the study, a pragmatic adjustment was made. Since each lexical level consisted of 1,000 words, A1 and A2 were combined to represent K1-K2, including the most fundamental 2,000 words of English. This combined category was used to classify the basic vocabulary easily comprehensible to most learners. Genres with a predominance of K1-K2 vocabulary were classified as CEFR A1

and A2. Texts with a higher level of complexity, dominated by K3, K4, and K5 vocabulary, were categorised as CEFR B1, B2, and C1, respectively. The most advanced texts, containing words from K6 through K25 and off-list, were designated as CEFR C2. In this study, the CEFR was utilised to analyse song lyrics, determining which genres were suitable for learners at different CEFR levels.

To quantify the lexical variation in the vocabulary of the songs in each genre in this study, the TTR (Laufer & Nation, 1995, p. 310) was employed, calculated as follows:

$$\text{TTR} = (\text{Number of Types/Number of Tokens}) \times 100$$

In this formula, ‘Number of Types’ denotes the count of unique words in the text (excluding repetitions), while ‘Number of Tokens’ refers to the total word count. Since the number of tokens can vary with text length (Gayraud, 2000; Kyle et al., 2021), this analysis evaluated individual songs to determine maximum, minimum, and average TTR values rather than examining the entire corpus. In this study, a high TTR indicated diverse vocabulary and greater opportunities for learning new words, while a low TTR reflected frequent repetition and limited vocabulary.

Finally, lexical density was also assessed. The calculation of lexical density was conducted using the following formula (Laufer & Nation, 1995, p. 309):

$$\text{Lexical Density} = (\text{Number of Content Words/Number of Tokens}) \times 100$$

In this formula, ‘Number of Content Words’ refers to words that carry the primary meaning, including nouns, verbs, adjectives, and adverbs, while ‘Number of Tokens’ represents the total word count in the text. In this study, a genre characterised by high lexical density indicated intricate and detailed lyrics, making it suitable for learners with advanced English proficiency. Conversely, genres with low lexical density signified less detailed lyrics, which may be more appropriate for learners with lower levels of English proficiency.

3. RESULTS

3.1 Percentage Representation of Lexical Profiling in Nine English Music Genres: GSL, AWL, and OWL

The lexical profiling analysis in Figure 1 reveals that coverage of the GSL in various music genres ranges from 78.88% to 91.33%. The rankings, from the highest to the lowest, are: jazz (91.33%), rock (90.26%), alternative (90.22%), country and folk (90.20%), electronic (88.36%), pop (88.17%), children’s (87.65%), R&B/soul (85.54%), and hip-hop (78.88%). These findings indicate that common English vocabulary is prevalent across all genres, although the extent of its usage differs by genre. Jazz, with the highest GSL coverage, is particularly suitable for L2 learners at a beginner level due to its high proportion of frequently used words.

In contrast, the coverage of the AWL is notably low, ranging from 0.23% to 0.60%. The rankings for the AWL coverage are: hip-hop (0.60%), R&B/soul (0.55%), alternative (0.45%), pop (0.33%), electronic (0.33%), children’s (0.33%), country and folk (0.25%), jazz (0.24%), and rock (0.23%). This indicates a minimal presence of academic vocabulary across all music genres.

The OWL coverage ranges from 8.43% to 20.52%, with hip-hop (20.52%) ranking the highest, followed by R&B/soul (13.91%), children’s (12.02%), pop (11.50%), electronic (11.31%), country and folk (9.55%), rock (9.51%), alternative (9.33%), and jazz (8.43%). These results highlight a substantial range of specialised terminology, technical vocabulary, or slang across genres. In particular, hip-hop stands out as a rich source of advanced lexical items, making it especially suitable for L2 learners aiming to expand their English proficiency.

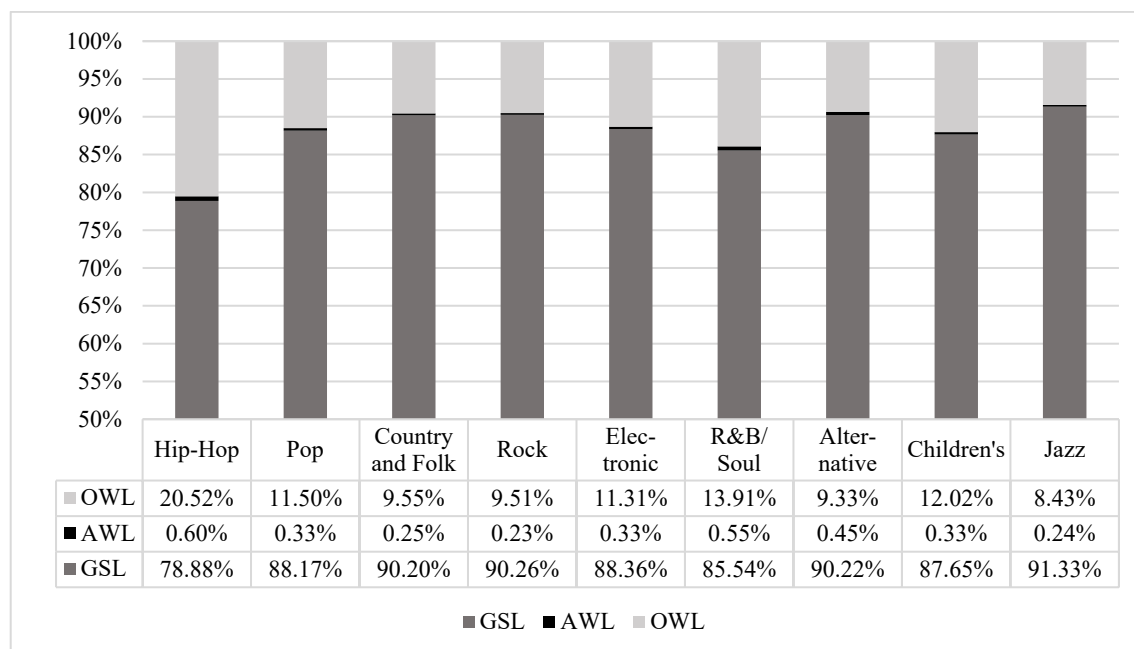


Figure 1. Lexical profiling of nine music genres.

The following Data 1 (D1) illustrates the GSL terms in regular text, the AWL terms in bold, and the OWL terms underlined. From (D1), which features a song in the hip-hop genre, it can be observed that if learners study vocabulary from this sample song, they might have the opportunity to encounter words from the GSL, AWL, and OWL. In particular, the OWL in hip-hop songs demonstrates a higher coverage rate compared to other genres, providing learners with a greater opportunity to acquire specialised or technical vocabulary.

- (D1) ‘Attention Deficit Hyperactivity Disorder, or better known as ADHD’
 ‘Is a **mental disorder** that **affects** an **individual’s** ability to **focus**’
 ‘Causing them to move around more frequently’
 ‘They may also have trouble controlling their **impulsive** behaviours’
 ‘One time for them pravin on my downfall (Yeah)’
 ‘Two times for the homies in the chow hall (Whoa)’
 ‘Three times for them hoes on the internet’
 (The hip-hop song entitled *Isis* by Joyner Lucas)

3.2 Distribution of High, Mid, and Low-Frequency Words in Nine English Music Genres: K1-K25 and Off-List Word Levels

Based on the analysis of nine music genres, the distribution of high-, mid-, and low-frequency words across lexical levels K1-K25 and off-list words is presented in Table 2. The analysis of the lexical level in different music genres reveals varying levels of suitability for language learners. It is noteworthy that the percentages for the lexical levels are presented to three decimal places due to their very small values, which would not be accurately displayed with only two decimal places.

Overall, country and folk songs (91.675%), alternative songs (90.118%), and jazz songs (90.067%) have the highest proportion of K-1 level vocabulary. This indicates that these genres use a high percentage of common, high-frequency words, making them highly accessible and ideal for beginners. In contrast, hip-hop has the highest proportion of ‘K-off’ words (7.266%), followed by R&B/soul and pop (5.12% and 4.56%, respectively), which include terms not found within the standard proficiency levels. This suggests that these genres use more complex and specialised vocabulary, making it more challenging and therefore well-suited for advanced learners looking to expand their language skills. Notably, while pop music relies heavily on

foundational vocabulary, it also contains a significant amount of K-off words (4.56%), which may be due to the use of slang or informal language.

Table 2. Lexical level of nine music genres (in percentages).

Level	Genre								
	Hip-hop	Pop	Country and Folk	Rock	Electronic	R&B/Soul	Alternative	Children's	Jazz
K-1	82.188	89.046	91.675	89.726	88.916	87.652	90.118	87.626	90.067
K-2	3.620	2.755	2.984	3.005	2.836	2.655	3.106	3.910	3.867
K-3	1.205	0.643	0.540*	0.818	0.638	0.919	1.004	0.814	0.809
K-4	1.237	0.698	0.968	0.967	0.829	0.932	0.762	0.912	1.108*
K-5	1.102	0.617	0.466	0.355*	0.557	0.649	0.518*	0.750	0.773
K-6	0.549	0.276	0.237	0.284	0.482	0.417	0.388	0.574	0.296
K-7	0.441	0.195	0.160	0.139	0.188	0.188	0.166	0.402	0.228
K-8	0.347	0.366	0.119	0.144	0.339	0.296	0.196	0.278*	0.126
K-9	0.387	0.344	0.104	0.227	0.310*	0.336	0.334	0.369	0.214
K-10	0.132	0.005	0.042	0.071	0.074	0.064	0.101	0.389	0.112
K-11	0.125	0.131*	0.076	0.036	0.159	0.072	0.055	0.096	0.069
K-12	0.102	0.030	0.033	0.102	0.103	0.107	0.032	0.367	0.107
K-13	0.098	0.027	0.041	0.052	0.031	0.073	0.063	0.253%	0.071
K-14	0.082	0.010	0.036	0.016	0.012	0.023	0.019	0.065	0.033
K-15	0.029	0.005	0.027	0.006	0.014	0.018	0.027	0.024	0.137**
K-16	0.023	0.006	0.021	0.003	0.018	0.011	0.014	0.022	0.016
K-17	0.039	0.011	0.012	0.001	0.002	0.014	0.010	0.029	0.047
K-18	0.021	0.005	0.011	0.001	0.005	0.008	0.008	0.002	0.011
K-19	0.037	0.007	0.003	0.035	0.007	0.032	0.035	0.013	0.025
K-20	0.007	0.006	0.005	0.003	0.007	0.005	-	0.018	0.038
K-21	0.009	0.002	-	0.003	0.031	0.013	0.002	0.085	0.016
K-22	0.010	-	0.001	0.003	0.005	0.006	0.006	0.005	0.003
K-23	0.025	0.004	0.005	0.003	0.004	0.009	-	0.007	0.005
K-24	0.002	-	-	-	0.001	0.002	-	0.002	0.016
K-25	0.004	-	-	-	0.001	0.003	0.003	-	0.005
K-off	7.266**	4.56**	2.423**	3.672**	4.25**	5.12**	2.792**	2.961**	1.799

Note. * = 95% lexical coverage ** = 98% lexical coverage

The analysis of the lexical coverage at 95% further reveals that country and folk songs correspond with K3 vocabulary, making them accessible for beginners. Jazz songs require K4 vocabulary, indicating a slightly higher complexity, while rock and alternative songs reach K5, making them suitable for intermediate learners. Children's songs utilise vocabulary at K8, likely due to specialised terminology. Electronic songs demand K9 and pop demands K11, indicating greater vocabulary diversity, and are therefore suitable for advanced learners. An interesting point regarding the lexical coverage analysis is that, among the nine music genres, eight require comprehension of off-list words to reach 98% coverage. This indicates that, aside from the jazz genre, most other music genres rely on a substantial proportion of low-frequency vocabulary, exceeding 2% on average. In the following section, the results of the lexical level analysis are presented according to lexical frequency bands, as illustrated in Figure 2.

Figure 2 illustrates that high-frequency words (K1-K3) are prevalent across all music genres, with the highest percentages found in country and folk (95.20%), jazz (94.74%), and alternative (94.23%). These genres are particularly suitable for beginners due to their reliance on frequently used vocabulary. Following these are rock (93.55%), pop (92.44%), electronic (92.39%), children's (92.35%), and R&B/soul (91.23%), all of which are appropriate for intermediate learners. Even though high-frequency words dominate across genres, notable distinctions in mid-frequency and low-frequency usage exist, with hip-hop standing out for its unique linguistic style. Hip-hop has a high-frequency word coverage rate (the combined count of words in K1-K3) of only 87.01, which is significantly lower than that of the other eight music genres. However, hip-hop surpasses the remaining genres in mid- and low-frequency word coverage, with rates of 4.06 (K4-K9) and 8.93 (K10 and above), respectively, which are notably

higher than those of the other genres. This indicates that hip-hop exhibits greater lexical complexity, making it more suitable for advanced learners.

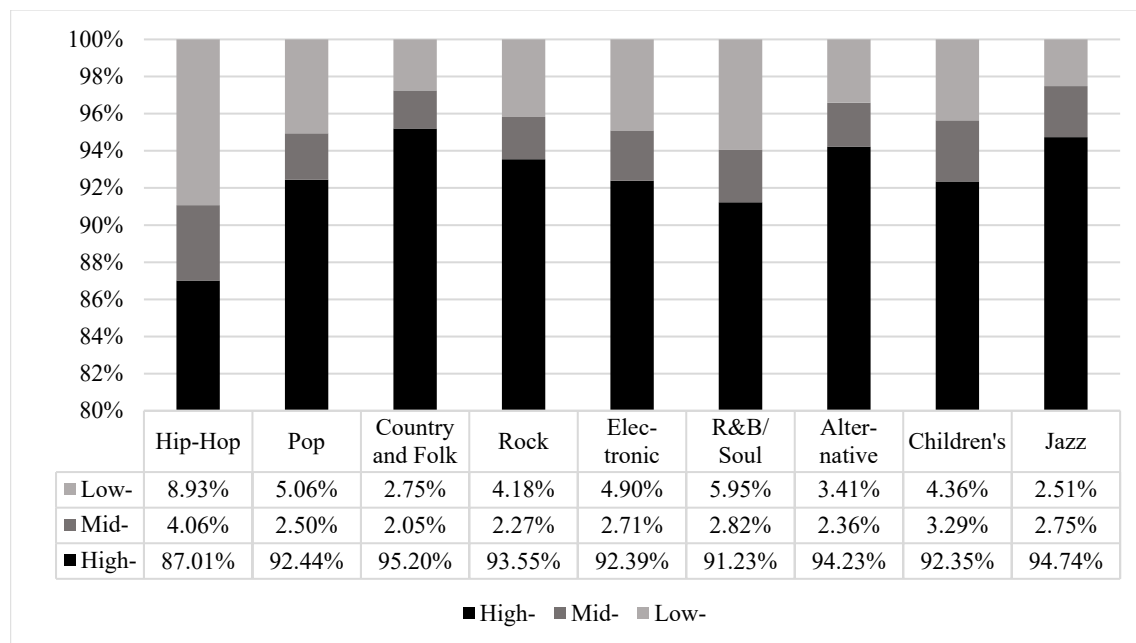


Figure 2. Lexical frequency groups of nine music genres.

In terms of mid-frequency words (K4-K9), hip-hop leads with 4.06%, followed by children at 3.29%. R&B/soul (2.82%), jazz (2.75%), electronic (2.71%), and pop (2.50%) exhibit moderate usage, while alternative (2.36%), rock (2.27%), and country and folk (2.05%) demonstrate lower percentages. Regarding low-frequency words (K10-K25+off-list), hip-hop also ranks as the highest at 8.93%, suggesting its appropriateness for advanced learners seeking to enhance their English skills through more complex vocabulary, including technical terms and slang. R&B/soul (5.95%), pop (5.06%), electronic (4.90%), children's (4.36%), and rock (4.18%) follow, indicating moderate low-frequency vocabulary usage, suitable for intermediate learners. Alternative (3.41%), country and folk (2.75%), and jazz (2.51%) show the least usage, suggesting these genres are better suited for beginners. Combining the percentages of mid-frequency and low-frequency words, hip-hop accounts for 12.99%, highlighting its distinctive linguistic features, characterised by a higher proportion of low-frequency vocabulary. Data 2 illustrates the high-frequency words in regular text, mid-frequency words in bold, and low-frequency words underlined.

(D2) 'To get ahead man you gotta make sacrifices'
 'F*ck it though that's how hungry my **appetite** is'
 'F*ck what you telling me, I sense the **jealousy**'
 'I got telepathy, I got the recipe'
 'I got the streets and I don't got no **felonies**'
 (The hip-hop song entitled *Sacrifices* by Big Sean)

3.3 CEFR Level Classification of Songs across Nine Music Genres

Figure 3 reveals that country and folk possess the highest percentage of vocabulary at A1-A2 (94.66%), indicating the use of basic, commonly encountered language. Conversely, hip-hop has the lowest percentage at 85.81%, suggesting a greater reliance on more complex vocabulary. At the B1 level, hip-hop leads with 1.21% of mid-frequency words, while country and folk have the lowest at 0.54%. For the B2 level, hip-hop maintains the highest percentage (1.24%), whereas pop has the lowest (0.70%). At the C1 level, hip-hop continues to dominate with 1.10%, highlighting the genre's use of highly complex words, while country and folk are at 0.47%.

Notably, at the C2 level, which caters to advanced learners, hip-hop shows the highest percentage of vocabulary at 10.64%, with country and folk at the lowest (3.36%).

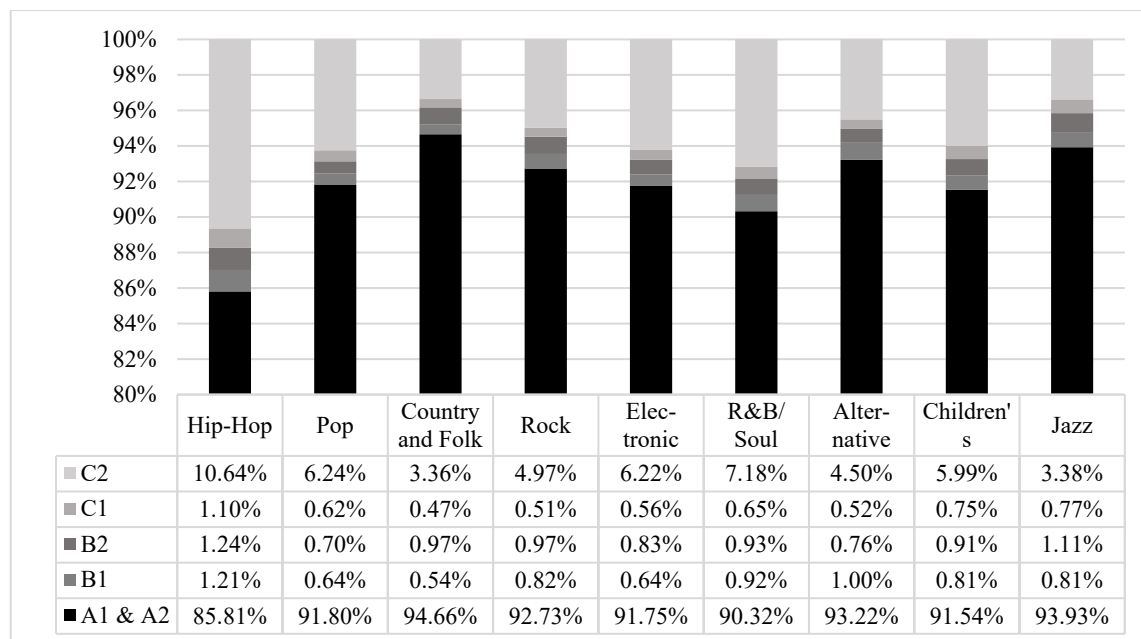


Figure 3. CEFR level of vocabulary across nine music genres.

Music genres can be categorised into beginner, intermediate, and advanced levels based on their relationship with CEFR levels. For beginners, country and folk are recommended due to their high percentage of high-frequency vocabulary. This is followed by jazz (93.93%), alternative (93.22%), and rock (92.73%), which predominantly feature A1-A2 vocabulary, suitable for basic practice. Intermediate learners may benefit from R&B/soul, children's, pop, and electronic songs, which contain vocabulary at B1 (average 0.64%-0.92%) and B2 (average 0.70%-0.93%) levels. Interestingly, children's songs display a relatively high percentage at the intermediate level, suggesting a more nuanced content than typically associated with the genre. For advanced learners, hip-hop is highly recommended, featuring complex vocabulary at C1 (1.10%) and C2 (10.64%) levels, ideal for those seeking to expand their language skills with less common terms. This categorisation aids in effectively utilising music as a tool for language learning according to CEFR levels. Data 3 and 4 illustrate the C2 in bold, italicised, and underlined, C1 in bold and underlined, B2 in bold and italicised, B1 in regular and underlined, and A2 and A1 in regular text.

(D3) 'It is not enough to take the one you love for granted'
 'You must **remind** her or she'll be inclined to say'
 'How do I know he loves me? (How does she know that you love her?)'
 '(How do you show her you love her?)'
 'How do I know he's mine?'
 '(How does she know that you really, really truly love her?)'
 'Well, does he leave a little note to tell you'
 'You are on his mind'
 (The children's song entitled *That's How You Know* by Amy Adams)

(D4) 'Count up the bands stickin' up. No **rubber** band is big enough'
 'Chain is so heavy, can't pick it up. Came with the **gang**, a myriad'
 'I seen your **whip**, **hilarious**. Where is my roof? **Mysterious**'
 'My **whip** from Fast & **Furious**. Curtains with yellow **interior**'
 (The hip-hop song entitled *Motley Crew* by Post Malone)

3.4 Lexical Variation in Nine English Music Genres: TTR Analysis

The TTR values range from the highest to the lowest as follows: jazz (43.44%), children’s (39.73%), hip-hop (34.49%), country and folk (33.42%), alternative (33.08%), rock (31.56%), R&B/soul (30.20%), pop (29.42%), and electronic (26.79%). Jazz and children’s genres display the highest lexical variation, making them ideal for learners seeking a wide range of vocabulary. In contrast, hip-hop, country, folk, and alternative exhibit moderate lexical variation, suitable for intermediate learners. The lowest lexical variation is found in pop and electronic genres, which are more appropriate for beginner learners due to their simpler vocabulary. The findings are presented in Figure 4.

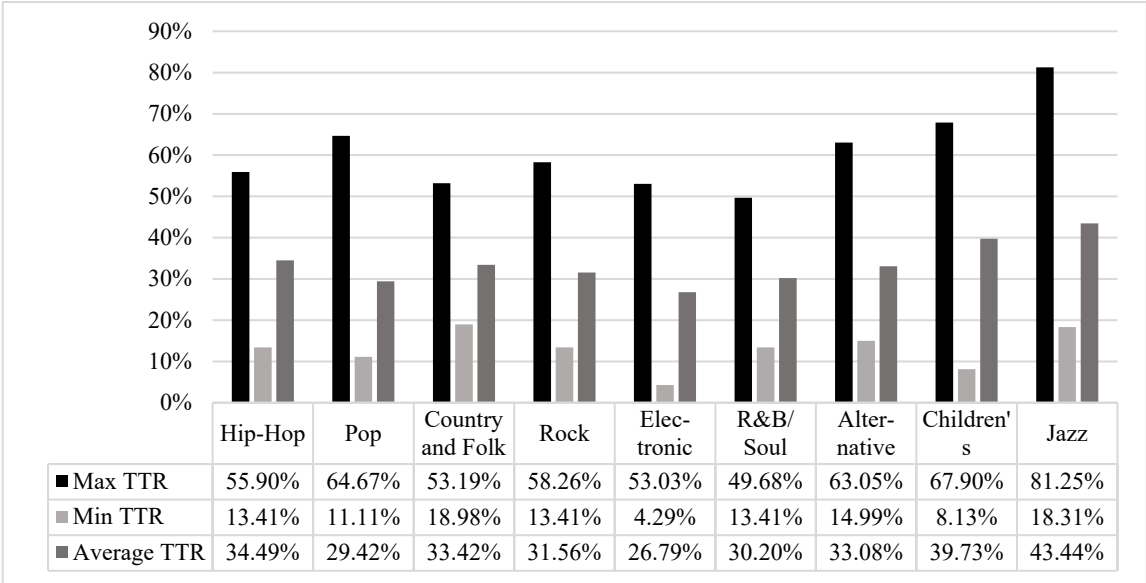


Figure 4. Lexical variation of nine music genres.

The use of diverse vocabulary reflects different methods of expressing content and emotion across genres. Genres like jazz, with high lexical variation, employ creative word usage to convey complex musical ideas. On the other hand, genres, such as pop and electronic, rely on straightforward terms for clarity and precision in communication. These findings highlight how vocabulary diversity enhances communicative richness and reflects each genre’s unique characteristics. Data 5 illustrating synonymous terms are presented with the first word in bold, the second in italics, and the third underlined.

(D5) ‘And let me sing for ever more. You are all I long for’
‘All I **worship** and *adore*’
‘In other words, please be true. In other words, I love you.’
(The jazz song entitled *Fly Me to the Moon* by Frank Sinatra & Count Basie)

3.5 Lexical Density Ratio in Nine English Music Genres

The analysis of the lexical density in the studied music genres highlights the proportion of content words in lyrics. A higher lexical density reflects a greater use of content words and fewer function words. The results, from the highest to the lowest, are: hip-hop (52.88%), children’s (51.80%), R&B/soul (50.25%), pop (49.87%), electronic (49.70%), rock (49.26%), alternative (48.61%), country and folk (48.27%), and jazz (48.23%). Hip-hop, with the highest lexical density, is well-suited for L2 learners aiming to acquire a large number of new words, as it contains a higher ratio of content words compared to function words. This enables learners to expand their vocabulary more effectively. Children’s and R&B/soul also exhibit high lexical density, while pop, electronic, and rock are more appropriate for intermediate learners.

Alternatively, country, folk, and jazz show the lowest lexical density, making them suitable for beginners due to their greater use of function words, as illustrated in Figure 5.

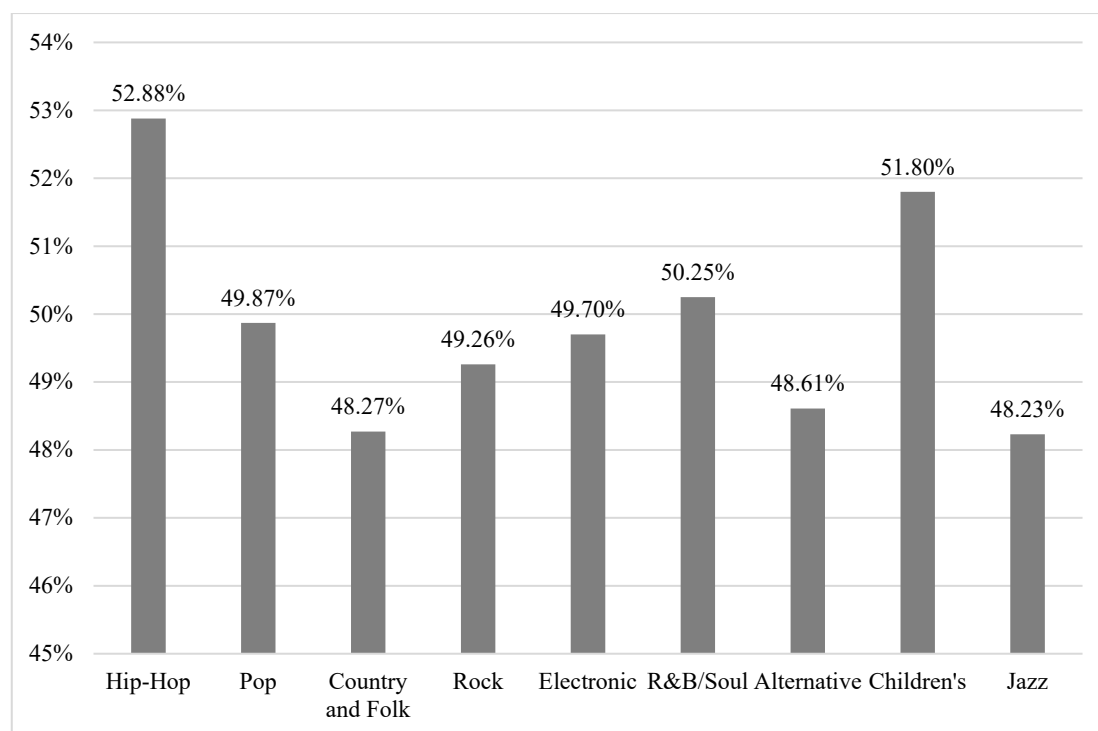


Figure 5. Lexical density of vocabulary across nine music genres.

When selecting music for language learning, it is also important to consider both lexical variation and density. Jazz, for example, has the highest lexical variation but the lowest density, meaning it uses fewer content words; however, these words are often diverse and non-repetitive, providing learners with greater opportunities to encounter new vocabulary. Hip-hop, in contrast, has the highest lexical density but lower variation, with frequent repetition of content words. This suggests that educators should balance both aspects when selecting music genres appropriate for learners' levels. Examples of word usage in lyrics are provided below, with function words in regular type and content words in bold.

(D6) 'King of my **city**, king of my **country**, king of my **homeland**'
 'King of the **filthy**, king of the **fallen**, we living again'
 'King of the **shooters**, **looters**, **boosters**, and **ghettos poppin**'
 'King of the **past**, **present**, **future**, my **ancestors watching**'
 (The hip-hop song entitled *Black Panther* by Kendrick Lamar)

(D7) 'Ain't got no home, ain't got no **shoes**'
 'Ain't got no **money**, ain't got no **class**'
 'Ain't got no **friends**, ain't got no **schooling**'
 'Ain't got no **wear**, ain't got no **job**'
 (The jazz song entitled *Ain't Got No, I Got Life* by Nina Simone)

4. DISCUSSION

Based on the lexical profiling analysis, jazz shows the highest GSL coverage at 91.33%, followed by rock (90.26%), alternative (90.22%), and country and folk (90.20%). This indicates that general vocabulary plays a significant role in these genres, making them suitable for L2 learners aiming to develop basic English language skills. Hip-hop has the lowest GSL coverage at 78.88%, below the expected 80% (Nation & Waring, 1997), suggesting that the use of general

vocabulary in this genre is relatively low. Thus, hip-hop might be less suitable for L2 learners seeking to quickly develop fundamental vocabulary understanding. Regarding AWL coverage, hip-hop has the highest rank at 0.60%, followed by R&B/soul (0.55%) and alternative (0.45%). Rock shows the lowest AWL usage at 0.23%. The low AWL coverage across genres reflects the minimal use of academic vocabulary in music. These findings are consistent with other research analysing non-academic texts. For instance, [Laosirattanachai et al. \(2025\)](#) found that the AWL accounted for only about 3.77% of the words in TED Talks, while [Loesnophaimongkhon et al. \(2025\)](#) reported that news articles contained only about 5.81% AWL words. These data indicate that the AWL vocabulary plays a limited role in music and is more relevant in academic contexts. Regarding OWL coverage, which includes specialised terms, slang, and words not found in the GSL and AWL, hip-hop leads with 20.52%, followed by R&B/soul (13.91%) and children's music (12.02%). Jazz has the lowest OWL coverage at 8.43%, making hip-hop more suitable for advanced learners who aim to advance their English language skills to a higher level and are interested in learning less common and more complex vocabulary. These findings suggest that L2 learners should select music based on their language proficiency level: jazz for beginners, focusing on basic vocabulary, and hip-hop for advanced learners seeking to expand their English skills with more specialised and challenging terms.

With respect to lexical levels, the findings show that high-frequency vocabulary (K1-K3) is most dominant in country and folk (95.20%), jazz (94.74%), and alternative (94.23%). The strong presence of such words shows that these genres are more accessible than others. Conversely, hip-hop uses the fewest high-frequency words (87.01%), indicating the genre's complexity and higher demands on L2 learners. In terms of mid-frequency words (K4-K9), hip-hop exhibited the highest usage (4.06%), followed by children's music (3.29%) and R&B/soul (2.82%). The use of mid-frequency vocabulary elevates the linguistic sophistication of the lyrics, making the language more nuanced and stylistically refined than communication that relies solely on high-frequency words. These findings are in line with [Meebangsai et al. \(2023\)](#), who highlighted the importance of mid-frequency vocabulary in enhancing content quality. For low-frequency words (K10-K25 and off-list), which are often specialised vocabulary and slang ([Schmitt & Schmitt, 2020](#)), the hip-hop genre also ranked the highest, with 8.93% use of less frequent words. This suggests that the high proportion of low-frequency words in hip-hop likely arises from the extensive use of technical terms and colloquial expressions in its lyrics. This makes hip-hop particularly suitable for advanced English learners seeking to improve higher-level language skills. Classroom instruction rarely covers vocabulary beyond K3 ([Stoffelsma, 2019](#)), whereas song lyrics, including hip-hop, span a wider lexical range, making them effective for vocabulary learning. Hip-hop's combined percentage of mid-frequency, low-frequency, and off-list words was 12.99%, emphasising the genre's complexity. This blend of vocabulary contributes to the depth of hip-hop lyrics, though it may challenge beginner learners. In contrast, country and folk music exhibit the lowest combined percentage (4.80%), suggesting simpler and more familiar vocabulary, suitable for novice L2 learners.

The analysis of the lexical coverage further reveals that country and folk achieved 95% coverage with a smaller vocabulary range at K3. This is followed by jazz at K4, rock and alternative at K5, children's at K8, electronic at K9, and pop at K11, indicating that a larger vocabulary range is necessary. For a deeper understanding of a song's content, 98% lexical coverage is required. In this case, the jazz genre allows for deeper comprehension with a vocabulary range at the K15 level. Conversely, other genres require vocabulary knowledge at the off-list level in order to achieve a thorough understanding of the lyrics. This reflects the proposal of [Schmitt et al. \(2011\)](#), which suggested that a broader vocabulary range enhances comprehension of content.

For CEFR levels, the analysis reveals that all genres heavily use A1-A2 level vocabulary, with country and folk having the highest usage. This suggests that country and folk (94.66%), jazz (93.93%), and alternative (93.22%) genres are suitable for beginners, as the vocabulary consists of basic and everyday words. This is consistent with [Nguansumangk \(2020\)](#), who found that most songs primarily used A1-level vocabulary, making them ideal for developing fundamental language skills. In contrast, hip-hop features more complex vocabulary, with notable

percentages at the C1 (1.10%) and C2 (10.64%) levels. This indicates the hip-hop genre's linguistic complexity, making it more suitable for advanced learners aiming to develop specialised vocabulary. Beginners (A1-A2) can benefit from listening to genres, such as country and folk, jazz, alternative, and rock, as these genres predominantly use basic vocabulary. Intermediate learners (B1-B2) should opt for genres like R&B/soul, children's, pop, and electronic, which balance high- and low-frequency vocabulary. Although often associated with younger audiences, children's music can also support mid-level language skills due to its nuanced vocabulary. For advanced learners (C1-C2), hip-hop is ideal, given its complex vocabulary, especially at the C2 level. The CEFR analysis in this study reveals both similarities and differences when compared with research examining other media, such as news (Loesnopchaimongkhon et al., 2025) and TED Talks (Laosrirattanachai et al., 2025). Across all three, the majority of vocabulary falls within the A1 and A2 levels, reflecting a shared intention to ensure accessibility for broad audiences. However, a striking difference appears at the C2 level, where songs average a markedly higher proportion (5.84%) compared with news (3.62%) and TED Talks (1.96%). This lexical complexity may stem from artistic creativity, stylistic expression, and the desire to craft linguistically memorable or salient lyrics that enhance retention, thereby introducing less frequent words.

Lexical variation analysis using the TTR reveals notable differences in vocabulary diversity across music genres, with an average TTR of 33.57%. Higher TTR values indicate a broader vocabulary and enhanced text quality (Jarvis, 2013; Johansson, 2008), while lower values reflect repetitive, limited vocabulary. Jazz (43.44%) and children's music (39.73%) exhibit the highest TTRs, reflecting rich and varied vocabulary beneficial for language learners, either for complex emotional expression or educational and narrative purposes. Genres with moderate TTRs, including hip-hop (34.49%), country and folk (33.42%), alternative (33.08%), rock (31.56%), and R&B/soul (30.20%), offer a balanced range suitable for intermediate learners. Hip-hop uniquely combines moderate TTR with repeated use of rare, advanced C1-C2 vocabulary, likely due to slang and specialised terms in memorable hooks. Pop (29.42%) and electronic (26.79%) display the lowest TTRs, favouring repetitive, straightforward language suited to beginners. The analysis of pop music, in particular, is consistent with Parada-Cabaleiro et al. (2024), who found that pop lyrics have become simpler and easier to comprehend over time in terms of vocabulary variation. Variation in TTR reflects differences in stylistic conventions, audience targeting, and lyrical composition, with some genres prioritising repetition for memorability and others encouraging diverse lexical choices for expressive or narrative purposes.

For the lexical density analysis, the results indicate that hip-hop possesses the highest lexical density at 52.88%, indicating a rich proportion of content words, making it an excellent resource for learners seeking to acquire meaningful vocabulary. Such high-density content can provide advanced learners with detailed and varied information. Similarly, children's and R&B/soul genres also display substantial lexical densities at 51.80% and 50.25%, respectively, making them suitable for learners wishing to expand their vocabulary. In contrast, pop, electronic, and rock genres have lower lexical densities of 49.87%, 49.70%, and 49.26%, respectively, suggesting a higher reliance on function words. Therefore, these genres are well-suited for intermediate learners. In contrast, the alternative, country, folk, and jazz genres, with TTR values of 48.61%, 48.27%, and 48.23%, respectively, rely more heavily on function words, making them more appropriate for beginners. Based on the research findings, the average lexical density of song lyrics is 49.88%. This value can be discussed in terms of its consistency, differences, and distinctiveness when compared to other types of text. The lexical density of song lyrics, at 49.88%, falls between the figures for written and spoken language as observed by other researchers. Compared with the lexical density of written language, which is around 58% (Meebangsai et al., 2023), song lyrics show a lower density. This is consistent with the fact that song lyrics, although written, are designed to be performed and heard. They often incorporate repetition and simpler syntax to be more accessible and memorable, which naturally lowers the proportion of lexical density. Conversely, the value is higher than the figure mentioned by Ure (1971), who stated that spoken language typically has a lexical density below 40%. This difference highlights the unique nature of song lyrics. Songs are not purely spoken. They are a

form of artistic expression, carefully crafted with a specific message or narrative. While they may share some characteristics with spoken language (e.g., direct address, colloquialisms), the creative and deliberate choice of words elevates their lexical density above that of casual conversation. The lexical density of song lyrics positions them as a hybrid form between written and spoken language. Their moderate density reflects their dual role as literary texts and performance art, with vocabulary carefully balancing descriptive, emotive, and narrative words.

5. CONCLUSION

This study analysed the vocabulary of nine music genres, revealing significant differences in their lexical characteristics. The results consistently show that jazz, country, folk, and alternative music genres are rich in high-frequency vocabulary, as indicated by their high coverage of the GSL and frequent use of high-frequency words. This makes them potentially suitable for beginners, as their vocabulary is largely in line with the A1-A2 CEFR levels. These genres also tend to have a high degree of lexical variation, particularly jazz, which suggests a rich and diverse vocabulary for language learners to explore. In contrast, hip-hop stands out as the most lexically complex genre. It exhibits the lowest GSL coverage but the highest usage of the OWL and a greater proportion of mid- and low-frequency words. This is further supported by the CEFR analysis, which shows that hip-hop contains the highest concentration of C1-C2 level vocabulary. Additionally, hip-hop, along with children's songs and R&B/soul, showed the highest lexical density, indicating a higher ratio of content words to function words. While this complexity may pose a challenge for beginners, it presents a valuable resource for advanced learners seeking to expand their vocabulary.

Selecting music for English language learning should consider vocabulary from multiple perspectives to be appropriate for the learner's proficiency level and educational goals. This research highlights the significant role of music as a tool for learning English and encourages careful consideration in selecting music that is appropriate for the learner's language proficiency and learning objectives. Educators and learners can use this information to improve teaching and learning processes more effectively.

The study is limited by the possibility that the selected music genres do not represent all possible types. Additionally, the analysis may be affected by the length of the lyrics, the context of the songs, and the writing styles of the artists. The final limitation is the inability to provide a complete list of vocabulary from the entire song corpus, due to its substantial size. Future research should expand the range of genres analysed and explore vocabulary in broader contexts in order to enhance instructional design and better support L2 learners' needs.

REFERENCES

- Anthony, L. (2024). *AntWordProfiler* (Version 2.2.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/AntWordProfiler>
- Cobb, T. (2022). *VocabProfile*. [Computer program]. <http://www.lexutor.ca/vp/>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, XII(2), 65-78. <https://doi.org/10.3917/rfla.122.0065>
- Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Blackwell Publishing.
- Deja, J. A., Blanquera, K., Carabeo, C. E., & Copiaco, J. R. (2016). Genre classification of OPM songs through the use of musical features. *Theory and Practice of Computation*, 2016, 77-88. https://doi.org/10.1142/9789814730464_0007

- Early, G., & Monson, I. (2019). Why jazz still matters. *Daedalus*, 148(2), 5-12. https://doi.org/10.1162/daed_a_01738
- Foleiss, J. H., & Tavares, T. F. (2019). Random projections of mel-spectrograms as low-level features for automatic music genre classification. *arXiv*. <https://doi.org/10.48550/arxiv.1911.04660>
- Gayraud, F. (2000). *Le développement de la différenciation oral/écrit vu à travers le lexique* [The development of oral and written differentiation as reflected in the lexicon] [Master's thesis, Université Lumière Lyon 2]. HAL-SHS. <https://shs.hal.science/tel-05248927>
- Ito, Y. (2019). Developing a reading test on CEFR-J. *The Journal of Kanda University of International Studies*, 31, 45-57. <https://doi.org/10.69236/0000001635>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1), 87-106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 53, 61-79. <https://journals.lub.lu.se/LWPL/article/view/2273>
- Kao, T., & Oxford, R. L. (2014). Learning language through music: A strategy for building inspiration and motivation. *System*, 43, 114-120. <https://doi.org/10.1016/j.system.2014.01.003>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Krishnan, P. D., & Yunus, M. M. (2019). Blended CEFR in enhancing vocabulary among low proficiency students. *Arab World English Journal (AWEJ) Special Issue on CALL*, 5, 141-153. <https://dx.doi.org/10.24093/awej/call5.11>
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154-170. <https://doi.org/10.1080/15434303.2020.1844205>
- Laosrirattanachai, P., Baothong, C., Laijud, K., & Laosrirattanachai, P. (2025). Dissecting vocabulary usage across TED talks topics: A corpus-based perspective. *LEARN Journal: Language Education and Acquisition Research Network*, 18(2), 943-973. <https://doi.org/10.70730/XUFI2893>
- Laosrirattanachai, P., & Laosrirattanachai, P. (2021). Applying lexical profiling to construct technical word lists for Thai tourist guides. *PASAA*, 62(1), 61-91. <https://digital.car.chula.ac.th/pasaa/vol62/iss1/3>
- Laosrirattanachai, P., & Laosrirattanachai, P. (2023). Analysis of vocabulary use and move structures of the World Health Organization emergencies press conferences on Coronavirus Disease: A corpus-based investigation. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), 121-146. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/263436>
- Laosrirattanachai, P., & Laosrirattanachai, P. (2025). Tracing tourism business research trends in Scopus-indexed journals using corpus-based and judgement-based approaches. *Humanities, Arts and Social Sciences Studies*, 25(1), 32-53. <https://doi.org/10.69598/hasss.25.1.268122>
- Laosrirattanachai, P., & Ruangjaroon, S. (2020). The word lists of hospitality service review construction. *Journal of Studies in the English Language*, 15(1), 107-158. <https://so04.tci-thaijo.org/index.php/jsel/article/view/211119>
- Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 50-86. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/248677>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Liu, C. Y. (2021). Examining the implementation of academic vocabulary, lexical density, and speech rate features on OpenCourseWare and MOOC lectures. *Interactive Learning Environments*, 31(8), 4924-4939. <https://doi.org/10.1080/10494820.2021.1987274>

- Loesnopchaimongkhon, R., Phrommopakorn, C., Chernchom, P., & Laosrirattanachai, P. (2025). A corpus-based analysis of lexical characteristics across English news categories for L2 pedagogical use. *Journal of Studies in the English Language*, 20(2), 36-65. <https://so04.tci-thaijo.org/index.php/jsel/article/view/279647>
- Meebangsai, D., Pongtin, P., Kitipoontanakorn, P., & Laosrirattanachai, P. (2023). Investigating proficiency of academic English in student writing: A comparative case study on vocabulary utilization in student research article writing vis-à-vis national and international research. *PASAA*, 67(1), 66-100. <https://doi.org/10.58837/CHULA.PASAA.67.1.3>
- Na Ayutthaya, J. A., Kunthonjinda, K., Somwang, K., & Laosrirattanachai, P. (2022). Making beverage service word list for English for Specific Purposes classroom. *reFLections*, 29(2), 325-343. <https://doi.org/10.61508/refl.v29i2.259524>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company.
- Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Nation, I. S. P., & Crabbe, D. (1991). A survival language learning syllabus for foreign travel. *System*, 19(3), 191-201. [https://doi.org/10.1016/0346-251X\(91\)90044-P](https://doi.org/10.1016/0346-251X(91)90044-P)
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.
- Nguansumangk, N. (2020). *A systematic review of songs used in research studies in English language learning* [Master's thesis, Thammasat University]. TU E-Thesis Archive Library. https://ethesisarchive.library.tu.ac.th/thesis/2020/TU_2020_6221042192_14696_15763.pdf
- O'Keeffe, A., & McCarthy, M. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3-13). Routledge. <https://doi.org/10.4324/9780203856949-2>
- Parada-Cabaleiro, E., Mayerl, M., Brandl, S., Skowron, M., Schedl, M., Lex, E., & Zangerle, E. (2024). Song lyrics have become simpler and more repetitive over the last five decades. *Scientific Reports*, 14. 5531. <https://doi.org/10.1038/s41598-024-55742-x>
- Sadiqzade, Z. (2025). Melodies in the classroom: Using songs to enhance English language learning. *Acta Globalis Humanitatis et Linguarum*, 2(3), 126-133. <https://doi.org/10.69760/aghel.0250020018>
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching* (2nd ed.). Cambridge University Press.
- Schreiber, H. (2015). Improving genre annotations for the million song dataset. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR* (pp. 241-247). MIREX. <https://doi.org/10.5281/zenodo.1414759>
- Schultz, N. C., & Huron, D. (2015). Catching the lyrics: Intelligibility in twelve song genres. *Music Perception*, 32(5), 470-483. <https://doi.org/10.1525/mp.2015.32.5.470>
- Spicher, L., & Sweeney, F. (2007). Folk music in the L2 classroom: Development of native-like pronunciation through prosodic engagement strategies. *Connections*, 1, 35-48. <https://digitalcommons.stmarys-ca.edu/school-liberal-arts-faculty-works/2842>
- Stoffelsma, L. (2019). English vocabulary exposure in South African township schools: Pitfalls and opportunities. *Reading & Writing*, 10(1), a209. <https://doi.org/10.4102/rw.v10i1.209>
- Sukirmiyadi, A. (2017). Teaching English to young learners through Indonesian-translated songs. *Journal of Physics: Conference Series*, 953, 012236. <https://doi.org/10.1088/1742-6596/953/1/012236>

- Tasnim, Z. (2022). Songs for EFL/ESL class: How to teach listening skills through music. *MEXTESOL Journal*, 46(3), 1-12. <http://dx.doi.org/10.61871/mj.v46n3-16>
- Tegge, F. (2018). Pop songs in the classroom: Time-filler or teaching tool? *ELT Journal*, 72(3), 274-284. <https://doi.org/10.1093/elt/ccx071>
- Thornbury, S., & Slade, D. (2006). *Conversation: From description to pedagogy*. Cambridge University Press.
- Tomczak, E., & Lew, R. (2019). The song of words: Teaching multi-word units with songs. *3L: The Southeast Asian Journal of English Language Studies*, 25(4), 16-33. <https://doi.org/10.17576/3L-2019-2504-02>
- Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, 39(3), 302-327. <https://doi.org/10.1093/applin/amw009>
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 193-302. <http://dx.doi.org/10.1109/TSA.2002.800560>
- Ure, J. (1971). Lexical density and category differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). Cambridge University Press.
- Webb, S. (Ed.). (2019). *The Routledge handbook of vocabulary studies* (1st ed.). Routledge. <https://doi.org/10.4324/9780429291586>
- West, M. (1953). *A general service list of English words*. Longman.
- Wundervald, B. D., & Zeviani, W. M. (2019). Machine learning and chord-based feature engineering for genre prediction in popular Brazilian music. *arXiv*. <https://doi.org/10.48550/arXiv.1902.03283>
- Zhang, Y., Baills, F., & Prieto, P. (2023). Singing songs facilitates L2 pronunciation and vocabulary learning: A study with Chinese adolescent ESL learners. *Languages*, 8(3), 219. <https://doi.org/10.3390/languages8030219>