ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

COMPARATIVE ANALYSIS OF THE NEW GENERAL SERVICE LIST AND BNC-COCA 25 WORD LISTS: IMPLICATIONS FOR VOCABULARY TEACHING AND ASSESSMENT IN ENGLISH LANGUAGE EDUCATION

Dr Muhammad Umer Azim
Department of English, Government College Township, Lahore, Pakistan
mumerazim@gmail.com
Zaheer Hussain
Department of English, National University of Modern Languages (NUML), Lahore Campus, Pakistan
zhussain@numl.edu.pk
Azhar Munir Bhatti
Department of English, Government Dyal Singh College, Lahore, Pakistan
azharmunir18@hotmail.com

Abstract
Vocabulary selection constitutes a foundational concern in English language teaching, as the principle of maximum output with minimum input guides pedagogical decision-making regarding which lexical items to prioritize for instruction (Nation, 2006). This study presents a comparative analysis of two influential high-frequency word lists: the New General Service List (NGSL; Browne et al., 2013) and the British National Corpus-Corpus of Contemporary American English 25 (BNC-COCA 25; Nation, 2012). Through vocabulary profiling using Lextutor (Cobb, 2022), the study examines the degree of overlap, divergence, and coverage potential of these lists to inform vocabulary teaching and assessment practices. The analysis reveals substantial convergence between the two lists, with only five NGSL words (bedroom, forever, online, shareholder, weekend) absent from the BNC-COCA first three levels due to their status as compound words. The BNC-COCA first three levels contain 612 additional word families beyond the NGSL, suggesting broader coverage potential. The findings indicate that both lists demonstrate high utility for general English instruction, though the BNC-COCA lists offer enhanced coverage for pedagogical applications. The study discusses implications for curriculum design, materials development, and learner assessment, arguing that corpus-derived word lists should serve as benchmarks for systematic vocabulary instruction rather than prescriptive constraints on lexical selection.

Keywords: New General Service List, BNC-COCA 25, vocabulary profiling, corpus linguistics, word frequency lists, English language teaching, lexical coverage

1. Introduction
The relationship between vocabulary knowledge and language proficiency has been established as one of the most robust findings in applied linguistics research (Hazenberg & Hulstijn, 1996; Hu & Nation, 2000; Schmitt, 2008). Hazenberg and Hulstijn (1996) demonstrated that a minimal receptive vocabulary of approximately 10,000 word families is required for non-native university students to comprehend academic texts effectively. Similarly, Hu and Nation (2000) found that unknown vocabulary density significantly predicts reading comprehension, with 98% coverage representing the threshold for adequate comprehension of unsimplified texts. These empirical findings render the selection of vocabulary for pedagogical purposes a matter of considerable practical and theoretical significance (Lessard, 2013). The principle guiding such selection should be the achievement of maximum linguistic coverage with minimum instructional investment, ensuring that learners acquire the most useful lexical items first before progressing to lower-frequency vocabulary (Nation, 2006, 2012).

The historical development of high-frequency word lists reflects evolving methodologies in corpus linguistics and pedagogical lexicography. Michael West's General Service List (GSL), published in 1953, represented a landmark achievement in vocabulary selection, combining objective frequency data with subjective pedagogical judgment to identify approximately 2,000 word families of general utility for English language learners (West, 1953). However, the GSL has faced criticism on multiple grounds: its reliance on a relatively small corpus of 2.5 million words, the dated nature of its source materials, and the absence of a clear operational definition of what constitutes a "word" for counting

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

purposes (Browne, 2014). These limitations motivated subsequent researchers to develop updated word lists utilizing larger, more representative corpora and more transparent unitization criteria (Nation, 2016).

The New General Service List (NGSL), announced by Browne et al. (2013) on the sixtieth anniversary of the GSL, addresses many of these limitations. Derived from a carefully selected 273-million-word subsection of the 2-billion-word Cambridge English Corpus (CEC), the NGSL comprises approximately 2,803 high-frequency words organized according to a modified lexeme approach (Browne et al., 2013). This unitization strategy counts headwords across all parts of speech and their inflected forms, differing from the word family approach of the GSL by excluding derived forms (Browne, 2014). The NGSL was designed to provide optimal coverage of general English texts with the fewest words possible, achieving approximately 92% text coverage—the highest reported coverage for any corpus-derived general English word list (Browne, 2014; Nation, 2016).

Concurrently, Nation (2012, 2016) developed the BNC-COCA word family lists through harmonization of the British National Corpus frequency data with Mark Davies' (2012) 450-million-word Corpus of Contemporary American English. The resulting BNC-COCA 25 lists divide vocabulary into 25 frequency levels of 1,000 word families each, accompanied by supplementary lists for proper nouns, marginal words, transparent compounds, acronyms, and foreign words (Nation, 2016). Unlike the NGSL, which was optimized for general English coverage, the BNC-COCA lists were designed primarily for learners of English as a foreign language, with particular attention to ensuring that survival vocabulary, academic vocabulary, and high-frequency spoken forms receive appropriate prioritization (Nation, 2016). Dang et al. (2020) compared the first 2,000 words of the BNC-COCA lists against other high-frequency lists using coverage, teacher perceptions, and learner knowledge criteria, concluding that the BNC-COCA 2000 represented the most suitable general high-frequency word list for L2 learners.

Despite the widespread adoption of both lists in pedagogical materials, course design, and assessment instruments (Cobb, 2022; Nation, 2016), relatively few studies have directly compared their content and coverage characteristics. Such comparison is essential for informed pedagogical decision-making, as teachers and materials developers must select word lists that align with their learners' needs, proficiency levels, and learning objectives (Benson & Madarbakus-Ring, 2022). This study addresses this gap by systematically comparing the NGSL and the first three levels of the BNC-COCA 25 through vocabulary profiling analysis, examining the degree of lexical overlap, identifying divergent items, and evaluating the relative coverage potential of each list for general English instruction.

2. Literature Review

2.1 Theoretical Foundations of Vocabulary List Development

The creation of pedagogical vocabulary lists rests upon several interconnected theoretical principles (Nation, 2006, 2016). First, the frequency principle posits that the most frequently occurring words in a language should be taught first, as they provide the greatest text coverage per item learned (Nation, 2006). Nation (2006) estimated that knowledge of the most frequent 2,000 word families provides approximately 80% coverage of general English texts, while mastery of the most frequent 8,000 families extends coverage to roughly 98%. This logarithmic relationship between vocabulary size and text coverage underscores the pedagogical efficiency of prioritizing high-frequency items (Schmitt, 2008).

Second, the range principle complements frequency by ensuring that selected words appear across diverse text types and registers rather than being concentrated in narrow domains (Nation, 2016). The BNC-COCA lists explicitly incorporate range as a selection criterion, requiring that words appear across multiple sub-corpora including spoken and written English, fiction and non-fiction, and American and British varieties (Nation, 2016). This dual criterion of frequency and range enhances the generalizability of word lists beyond any single corpus or genre (Cobb, 2022). Dang and Webb (2016) evaluated multiple high-frequency word lists using coverage of both spoken and

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

written texts, finding that the BNC-COCA lists performed well across modalities compared to alternative lists.

Third, the concept of word families versus alternative unitization schemes has generated substantial theoretical discussion (Bauer & Nation, 1993; Nation, 2016). Bauer and Nation (1993) proposed a hierarchical model of word family levels based on morphological transparency, ranging from inflectional variants (Level 2) through increasingly complex derivational patterns (Levels 3–6). The BNC-COCA lists employ Level 6 word families, encompassing all regular inflections and a broad array of derivational affixes (Nation, 2016). In contrast, the NGSL utilizes modified lexemes, which include a headword across all parts of speech plus inflected forms but exclude derived forms (Browne, 2014). Dang et al. (2020) compared the first 2,000 words of the BNC-COCA lists against other high-frequency lists using coverage, teacher perceptions, and learner knowledge criteria, concluding that the BNC-COCA 2000 represented the most suitable general high-frequency word list for L2 learners.

2.2 Corpus Composition and List Validity

The validity of corpus-derived word lists depends fundamentally upon the representativeness of their source corpora (Nation, 2016; Schmitt, 2008). The NGSL draws from the Cambridge English Corpus, a 2-billion-word collection encompassing learner data, fiction, journals, magazines, non-fiction, radio, spoken discourse, documents, and television (Browne et al., 2013). The 273-million-word subsection selected for NGSL development was specifically designed to balance these registers appropriately (Browne, 2014). The BNC-COCA lists, by contrast, derive from harmonization of two major corpora: the 100-million-word British National Corpus and the 450-million-word Corpus of Contemporary American English (Davies, 2012; Nation, 2016).

For the first two 1,000-word-family lists, Nation (2016) constructed a special 10-million-word corpus with 60% spoken content (including movies and television) and 40% written content (including children's texts and fiction), deliberately overweighting spoken registers to compensate for the written bias of the BNC. This difference in corpus composition has significant implications for list content. The BNC-COCA's enhanced spoken representation ensures inclusion of high-frequency conversational items (e.g., alright, hello, dad, bye) that might be underrepresented in more formally oriented corpora (Nation, 2016). The NGSL's reliance on the Cambridge English Corpus, with its substantial learner data component, may differently weight items characteristic of learner interlanguage production (Browne, 2014). These compositional differences, while both defensible on pedagogical grounds, contribute to divergences in list content that warrant empirical investigation (Cobb, 2022).

2.3 Vocabulary Profiling as Analytical Methodology

Vocabulary profiling, as operationalized through tools such as Lextutor's VocabProfile and AntWordProfiler, provides a systematic methodology for comparing word lists and analyzing text lexical composition (Cobb, 2022). These tools classify vocabulary into frequency bands (typically K1, K2, K3, etc., representing the first, second, and third 1,000 word families) and calculate the percentage of text tokens covered by each band (Cobb, 2022). Such analysis enables researchers to determine the lexical demands placed on readers by particular texts, to evaluate the recycling of vocabulary across text sequences, and to compare the coverage potential of alternative word lists (Cobb, 2022; Nation, 2016).

The validity of vocabulary profiling depends upon the word lists employed as reference standards (Nation, 2016). Cobb and Morris (2004) demonstrated that vocabulary profiles can predict the academic performance of TESL trainees, with significant correlations between program success and the proportion of academic and off-list vocabulary in entrance essays. Similarly, Benson and Madarbakus-Ring (2022) utilized the BNC-COCA 25 and JACET8000 lists to analyze textbook vocabulary load, finding that learners faced higher lexical demands when evaluated against the Japanese-specific list compared to the general BNC-COCA list. These applications underscore the

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

practical significance of word list selection for pedagogical assessment and materials evaluation (Nation, 2016).

3. Methodology

3.1 Research Objectives

This study pursues three primary objectives: first, to compare the lexical content of the NGSL and the first three levels of the BNC-COCA 25 to identify areas of convergence and divergence; second, to evaluate the relative coverage potential of these lists for general English texts; and third, to derive pedagogical implications for vocabulary teaching and assessment based on the comparative findings (Nation, 2016).

3.2 Data Sources

The NGSL version 1.01, comprising 2,803 headwords, was obtained from the official NGSL website (Browne et al., 2013). The BNC-COCA 25 word family lists, specifically the first three 1,000-word levels, were accessed through Paul Nation's repository at Victoria University of Wellington (Nation, 2012). The BNC-COCA lists were utilized in their standard format, with word families defined according to Bauer and Nation's (1993) Level 6 criteria, encompassing regular inflections and derivational affixes including un-, -ly, -er, -th, -able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, and re- (Nation, 2016).

3.3 Analytical Procedure

The comparative analysis was conducted using Lextutor's Vocabulary Profiler (VP-Compleat), a web-based tool that categorizes input vocabulary according to reference word lists and calculates frequency-based statistics (Cobb, 2022). The NGSL headwords were input as the target text, with the BNC-COCA 25 lists serving as the profiling standard. This procedure generated classification of NGSL items across BNC-COCA frequency levels (K1, K2, K3, K4–K5, K6–K7, and off-list categories), enabling precise identification of lexical overlap and divergence (Cobb, 2022; Nation, 2016).

The analytical procedure involved several stages. First, the complete NGSL was submitted to the profiler to establish overall categorization patterns (Browne et al., 2013). Second, items classified as off-list or beyond K3 were extracted for detailed examination to identify the nature of divergence (e.g., compound words, specialized terminology, or genuinely discrepant frequency rankings). Third, the cumulative coverage statistics were compared to evaluate the relative comprehensiveness of the BNC-COCA first three levels versus the complete NGSL (Nation, 2016).

4. Results

4.1 Overall Categorization of NGSL Items

The vocabulary profiling analysis categorized the 2,803 NGSL headwords across the BNC-COCA 25 frequency levels. The results demonstrate substantial convergence between the two lists, with the vast majority of NGSL items falling within the BNC-COCA first three 1,000-word levels. Table 1 presents the frequency distribution of NGSL items across BNC-COCA levels.

Table 1

Categorization of NGSL Items by BNC-COCA Frequency Level

Freq. Level	Families	Types	Tokens
K-1 :	951	1193	1193
K-2 :	746	885	885
K-3 :	662	715	715
K-5 :	3	3	3
K-6 :	1	1	1
K-7 :	1	1	1

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Off-List:	??	5	5
Total (unrounded)	2364+?	2803 (100)	2803 (100)

The data reveal that 2,359 NGSL items (84.2%) fall within the BNC-COCA first three levels, distributed across K-1 (42.6%), K-2 (31.6%), and K-3 (25.5%). An additional five items appear at K-5, K-6, and K-7 levels, while only five items (0.2%) are classified as off-list in the BNC-COCA system.

4.2 Analysis of Off-List and Discrepant Items

The five NGSL items classified as off-list in BNC-COCA are all compound words: bedroom, forever, online, shareholder, and weekend. These items are excluded from the BNC-COCA base lists because the profiling system treats transparent compounds as separate category items rather than incorporating them into frequency-ranked base lists (Nation, 2016). This treatment reflects a principled decision regarding unitization rather than genuine lexical discrepancy—these compound words are unquestionably high-frequency items in English, but their morphological transparency renders them analytically distinct from simplex vocabulary in the BNC-COCA framework (Bauer & Nation, 1993; Nation, 2016).

Beyond the off-list items, five additional NGSL words appear beyond the third 1,000-word level in BNC-COCA. The word accord appears at K-6, while grammar, nowadays, noun, and verb appear at K-5 and K-7 levels. The elevated BNC-COCA rankings for these items likely reflect register-specific distribution patterns (Nation, 2016). Grammar, noun, and verb, while fundamental to metalanguage and educational discourse, may achieve lower overall frequency in general corpora dominated by non-academic text types (Davies, 2012; Nation, 2016). The word nowadays, conversely, may be more characteristic of written registers than the spoken-heavy BNC-COCA first 2,000 (Nation, 2016). These discrepancies illustrate how corpus composition and register balance influence frequency rankings and, consequently, list composition (Cobb, 2022; Schmitt, 2008).

4.3 Coverage Comparison

The BNC-COCA first three levels encompass 3,000 word families, compared to the NGSL's approximately 2,800 modified lexemes. This numerical difference translates into differential coverage potential. The BNC-COCA first three levels contain 612 more word families than the NGSL (3,000 versus 2,369 families falling within the first three levels), suggesting broader lexical representation (Nation, 2016). This additional coverage derives from the BNC-COCA's explicit design for L2 learners, which incorporates pedagogically motivated adjustments such as including all number words and weekdays in early levels and ensuring spoken high-frequency forms receive appropriate prioritization (Nation, 2016).

The coverage analysis by individual NGSL levels reveals interesting distributional patterns. For NGSL Level 1 items (approximately 1,000 headwords), 72.1% fall within BNC-COCA K-1, 23.3% within K-2, and 4.5% within K-3, yielding 99.9% coverage within the first three BNC-COCA levels (Browne et al., 2013; Nation, 2016). Only the word weekend appears off-list, and accord appears at K-6. For NGSL Level 2, the distribution shifts: 32.8% K-1, 35.5% K-2, and 31.7% K-3, with bedroom and online appearing as off-list compounds. NGSL Level 3 shows further dispersion: 15.9% K-1, 37.6% K-2, and 46.0% K-3, with forever and shareholder as off-list compounds, and grammar, nowadays, noun, and verb appearing at K-5 and K-7 (Browne et al., 2013; Nation, 2016).

This progressive dispersion across NGSL levels suggests that the NGSL's internal frequency ranking diverges from BNC-COCA's ordering, particularly for lower-frequency items (Nation, 2016). While the highest-frequency NGSL words (Level 1) overwhelmingly align with BNC-COCA K-1, the NGSL Level 3 items are more evenly distributed across BNC-COCA K-2 and K-3. This pattern may reflect differences in corpus composition, with the Cambridge English Corpus and BNC-COCA assigning divergent frequency weights to mid-frequency items (Browne, 2014; Nation, 2016).

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

5. Discussion

5.1 Convergence and Complementarity

The finding that 99.8% of NGSL items appear within the BNC-COCA first three levels (when accounting for the compound word classification) indicates substantial convergence between these major pedagogical word lists. This convergence validates the underlying frequency principle that guides both lists: the highest-frequency words in English are sufficiently robust across corpora that different compilation methodologies yield largely overlapping results (Nation, 2006, 2016). Teachers and materials developers can therefore employ either list with reasonable confidence that they are addressing the core high-frequency vocabulary of English (Dang et al., 2020).

However, the complementarity between the lists should not be overlooked. The BNC-COCA's 612 additional word families within the first three levels represent significant expanded coverage that may benefit learners in diverse contexts (Nation, 2016). Nation's (2016) explicit design decisions—to overweight spoken registers, to include functional vocabulary (numbers, days, months) at early levels, and to ensure academic vocabulary receives appropriate prioritization—enhance the pedagogical utility of the BNC-COCA lists for general English courses. The NGSL's achievement of 92% text coverage with fewer items, conversely, demonstrates remarkable efficiency for contexts where rapid coverage acquisition is prioritized (Browne, 2014; Nation, 2016).

5.2 The Compound Word Issue

The treatment of compound words represents a systematic source of divergence between the NGSL and BNC-COCA lists. The BNC-COCA profiling system's exclusion of transparent compounds from base lists reflects a morphological analysis principle: compounds whose meanings are compositional (bedroom = bed + room) need not be learned as opaque lexical items (Bauer & Nation, 1993; Nation, 2016). However, pedagogical practice often treats high-frequency compounds as single learning units, particularly for lower-proficiency learners who may not yet possess the analytical capacity to decompose morphological structures (Schmitt, 2008). The NGSL's inclusion of these items acknowledges their functional utility in general English, while the BNC-COCA's exclusion prioritizes morphological systematicity (Browne, 2014; Nation, 2016).

This divergence has practical implications for materials development. Teachers utilizing BNC-COCA-based profiling should anticipate that transparent compounds will appear as off-list items and may require supplementary instruction (Nation, 2016). Conversely, NGSL-based materials automatically incorporate these compounds, potentially reducing the cognitive load for learners who benefit from holistic lexical acquisition (Browne, 2014).

5.3 Pedagogical Implications

The comparative findings support several pedagogical recommendations. First, the substantial overlap between lists suggests that institutional investment in either framework will yield comparable benefits for core vocabulary instruction (Dang et al., 2020). Schools and textbook publishers need not agonize over list selection, as both NGSL and BNC-COCA provide valid foundations for high-frequency vocabulary teaching (Nation, 2016).

Second, the BNC-COCA lists' broader coverage within the first three levels recommends them as benchmarks for comprehensive vocabulary assessment (Nation, 2016). The additional 612 word families provide finer gradation of learner knowledge and may reduce the proportion of off-list vocabulary encountered in general texts, thereby supporting more accurate proficiency measurement (Cobb, 2022; Nation, 2016).

Third, the identified divergent items—particularly the off-list compounds and the mid-frequency discrepancies (grammar, noun, verb, nowadays)—should inform supplementary vocabulary instruction (Schmitt, 2008). Teachers can anticipate that these items, while not universally prioritized across lists, represent genuine high-frequency vocabulary that learners will encounter in authentic language use (Nation, 2016).

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Fourth, the modified lexeme versus word family unitization difference between NGSL and BNC-COCA has implications for morphological instruction. The NGSL's exclusion of derived forms places greater burden on learners to recognize derivational patterns productively, while the BNC-COCA's inclusion of derived family members provides more explicit morphological scaffolding (Bauer & Nation, 1993; Nation, 2016). Materials developers should align their morphological teaching sequences with their chosen list's unitization principles (Schmitt, 2008).

6. Conclusion

This comparative analysis of the NGSL and BNC-COCA 25 word lists reveals substantial convergence alongside systematic divergences that reflect differing corpus compositions, unitization principles, and pedagogical design priorities (Browne, 2014; Nation, 2016). The finding that 99.8% of NGSL items align with BNC-COCA first three levels validates the robustness of frequency-based vocabulary selection across methodological variations. The BNC-COCA lists' broader coverage, achieved through enhanced spoken register representation and pedagogically motivated adjustments, recommends them as particularly suitable benchmarks for general English teaching and assessment in diverse L2 contexts (Dang et al., 2020; Nation, 2016).

The study's limitations include its exclusive focus on lexical content comparison without direct empirical validation through learner performance data or text coverage analysis (Cobb, 2022). Future research should examine how the identified divergences between lists affect actual reading comprehension, writing quality, and learner vocabulary growth trajectories (Schmitt, 2008). Additionally, the analysis was restricted to the first three levels of the BNC-COCA lists; extension to higher levels would illuminate how the lists diverge in mid- and low-frequency vocabulary, which becomes increasingly relevant for advanced learners (Nation, 2006, 2016).

The ongoing development of word list technology, including semantic annotation using large language models, promises to enhance the precision of vocabulary selection beyond current frequency-based approaches (Cobb, 2022). Such innovations may eventually address limitations of existing lists, including the treatment of polysemous words, metaphorical meanings, and register-specific usage patterns (Nation, 2016). Nevertheless, the present analysis confirms that corpus-derived word lists, whatever their specific methodological differences, provide empirically grounded foundations for systematic vocabulary instruction that far exceed intuition-based lexical selection (Dang & Webb, 2016; Nation, 2016). Teachers, materials developers, and assessment designers should embrace these tools as benchmarks for pedagogical decision-making while remaining attentive to the contextual needs of their specific learner populations (Schmitt, 2008).

References

Bauer, L., & Nation, I. S. P. (1993). Word families. International Journal of Lexicography, 6(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Benson, S., & Madarbakus-Ring, N. (2022). A comparison of textbook vocabulary load analysis. Vocabulary Learning and Instruction, 11(2), 1–15. https://doi.org/10.7820/vli.v11.2.benson

Browne, C. (2014). The New General Service List: Celebrating 60 years of vocabulary learning. The Language Teacher, 38(4), 12–16.

Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List (NGSL) [Data set]. http://www.newgeneralservicelist.org/

Cobb, T. (2022). Lextutor VocabProfile [Computer software]. https://www.lextutor.ca/vp/

Cobb, T., & Morris, L. (2004). Vocabulary profiles as predictors of the academic performance of TESL trainees. System, 32(1), 75–87. https://doi.org/10.1016/j.system.2003.04.002

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. ITL–International Journal of Applied Linguistics, 167(2), 132–158. https://doi.org/10.1075/itl.167.2.01dan

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. Language Teaching Research, 24(6), 741–763. https://doi.org/10.1177/1362168818805172

Davies, M. (2012). Corpus of Contemporary American English (COCA) [Data set]. https://www.english-corpora.org/coca/

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. Applied Linguistics, 17(2), 145–163. https://doi.org/10.1093/applin/17.2.145

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. Reading in a Foreign Language, 13(1), 403–430.

Lessard, G. (2013). Vocabulary and language learning: The role of word lists. In C. A. Chapelle (Ed.), The encyclopedia of applied linguistics (pp. 1–6). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1293

Mushtaq, H., Bhatti, A.M. & Yasmin, T. (2021). A Corpus Based Vocabulary Analysis of Intermediate Book 1 used in the Colleges of Punjab. *Competitive Linguistic Research Journal*, 2(1), 31-57.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? Canadian Modern Language Review, 63(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2012). The BNC/COCA word family lists [Data set]. Victoria University of Wellington. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, I. S. P. (2016). Making and using word lists for language learning and testing. John Benjamins. https://doi.org/10.1075/z.195

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. Language Teaching Research, 12(3), 329–363. https://doi.org/10.1177/1362168808089921

West, M. (1953). A general service list of English words. Longman, Green & Co.