

# General Word Lists: Overview and Evaluation

Dana Therova  
*The Open University*

## Abstract

Vocabulary learning is unarguably one of the sub-goals in every language classroom. The learning and teaching of vocabulary have been transformed by the development of general word lists, providing compilations of the most prevalent vocabulary items used in everyday contexts. These lists have made an invaluable contribution to the field of applied linguistics in terms of both research and pedagogy; they have assisted the learning, teaching and testing of vocabulary; and they have also been widely used in materials development and vocabulary research. However, if they are to be utilised effectively, it is important to understand the characteristics of these word lists. Thus, this article offers a review of the various general word lists presently available with the aim of assisting English as a Foreign Language (EFL)/English as a Second Language (ESL) practitioners in making informed decisions regarding the choice and utility of these word lists in their practices.

## 1. Introduction

Learning vocabulary is a complex process that requires learners to familiarise themselves not only with the form but also with the variety of meanings carried by a given lexical item (Brezina & Gablasova, 2015). The learning and teaching of vocabulary have been transformed by the development of word lists with their primary purpose to expose learners to the most frequently occurring lexical items in the contexts in which they need to operate. Such word lists can “guide both English teacher and student attention and efforts for both comprehension and production of English vocabulary” (Lessard-Clouston, 2013, p. 299). In addition to learning, “[w]ord lists lie at the heart of good vocabulary course design, the development of graded materials for extensive listening and extensive reading, research on vocabulary load, and vocabulary test development” (Nation, 2016, p. xi). Thus, the importance of word lists in research and pedagogy seems indisputable as they provide compilations of vocabulary in common use in various contexts, ranging from everyday English to more specialised contexts such as specific professions or academic disciplines. These various contexts in which word lists are used reflect Nation’s (2001) distinction between four kinds of vocabulary: high-frequency words, technical words, low-frequency words and academic words. High-frequency words occur in various uses of the language and have a large coverage of texts, while low-frequency words do not occur with high frequencies and cover only small proportions of texts. Technical words are prevalent in a specific subject or topic area, and academic words are frequently used in a wide range of academic texts (Nation, 2001). Given the importance of high-frequency vocabulary for L2 learners

of English, and thereby the role that general word lists can play in the learning and teaching of such vocabulary items, this article focuses on general word lists providing compilations of high-frequency vocabulary. It starts with the oldest compilation of general words, namely, West's (1953) General Service List of English Words, followed by two new General Service Lists (Brezina & Gablasova, 2015; Browne et al., 2013) created in response to West's List. Next, Nation's (2006, 2012) BNC2000 and British National Corpus (BNC)/Corpus of Contemporary American English (COCA) word lists are discussed. Finally, the most recent Essential Word List (Dang & Webb, 2016a) is focused on. The article concludes with an overview of the main characteristics of the reviewed word lists and their pedagogical implications.

## 2. General word lists

### 2.1. The General Service List of English Words

One of the most widely used general word lists is West's (1953) General Service List of English Words (GSL). What is commonly known as the GSL, however, is in fact a revised version of the 'Interim Report on Vocabulary Selection' (Faucett et al., 1936), which resulted from the aim to "simplify English for learners" by finding "the minimum number of words that could operate together in constructions capable of entering into the greatest variety of contexts" (West, 1953, p. v). The report contained a list of approximately 2,000 words "considered suitable as the basis of vocabulary for learning English as a foreign language" (West, 1953, p. vii). So as to target a specific linguistic need, which was seen in selecting core vocabulary items of general application relevant to foreign language instruction, the report was revised by Dr. Michael West, whereby the GSL was conceived. The discussion of the GSL first focusses on its characteristics and influence in both research and pedagogy, followed by its evaluation.

The GSL was built on the basis of a 5-million word corpus of written English using quantitative measures of word frequency and several qualitative criteria (including ease or difficulty of learning, necessity, stylistic level, and intensive and emotional words) (West, 1953, pp. ix–x). The resulting GSL contains the most frequently used 2,000 word families in English divided into the first and second most frequent 1,000 words listed alphabetically. The GSL was extensively revised by Paul Nation in the early 1990s, which resulted in numerous inflected and derived forms being added (e.g., *broader* and *broadly* under *broad*) together with numbers, months, days of the week and letters of the alphabet, and some compound forms (e.g., *broadcast*) being excluded (Stoeckel, 2019).

The GLS has been widely used in pedagogical practice, where it has served as a basis for introducing students to the most common English words. It has also been used as the non-academic baseline for the development of the Academic Word List (Coxhead, 2000). In addition, it has contributed to vocabulary research with a number of analytical tools having incorporated the GSL, including the Compleat Lexical Tutor (called VocabProfile) (Cobb, n.d.), the Range (Nation & Heatley, 1994) and AntWordProfiler (Anthony, 2013). However, despite its far-reaching influence in both research and pedagogy, the GSL has been subject to criticism, which is discussed next.

Inevitably, due to the changes that have occurred in the English language and culture since the GSL's creation in the 1950s, some of the words contained in the list are used less frequently and hence are of less relevance today (e.g., *coal*, *vessel*, *ornament*, *telegraph*, *milk-maid*, *cart*, *oar*, *merchant*, *mill*, *bless*, *grace*, *preach* and *shilling*), while some words in frequent use today (e.g., *computer*, *astronaut*, *internet* and *television*) are not included (Brezina & Gablasova, 2015; Browne, 2014; Paquot, 2010; Richards, 1974). Similarly, owing to the advances in corpus techniques, the size of the source corpus used for the GSL's compilation may be regarded as somewhat limited in its scope by modern standards.

A further potential weakness of the GSL can be seen in the fact that West did not use a systematic definition of what constituted a word: "No attempt has been made to be rigidly consistent in the method used for displaying the words: each word has been treated as a separate problem, and the sole aim has been clearness" (West, 1953, p. viii). Another criticism relates to the selection of the words included in the GSL, as noted by Richards (1974), who found inconsistencies with certain vocabulary items belonging to the semantic category of animals (e.g., *bear*, *cat* and *horse*) being included, but others (e.g., *tiger*, *lion* and *fox*) being excluded. Similarly, the category of occupations includes *doctor*, *teacher* and *nurse*, for instance, but *plumber* or *carpenter* is omitted. Another potential limitation of the GSL concerns its derivation from a corpus comprising written English, meaning that words common in spoken English are under-represented in the lists.

The use of word families as an organising principle has also been regarded as a limitation, particularly for pedagogical uses with lower level learners with limited word building skills. A word family refers to "a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (Bauer & Nation, 1993, p. 253). While the word family principle assumes that little effort is required for the recognition of other members of the word family once the base word is known, this presents the problem of the different meanings of the words belonging to the same word family. For example, the word family *impose* contains the following members: *impose* (headword), *imposed*, *imposes*, *imposing*, *imposition*. As word families do not distinguish between word classes, in the example above the adjectival and verbal uses of the form *imposing* (as in an *imposing* building versus stop *imposing* your beliefs on me) are merged. Among further examples are *please* and *unpleasantly* or *part* and *particle* (included in the same word family in West's GSL), clearly demonstrating the problem with assuming that the meanings of the words contained in the same word family will be easily deduced once the learners are familiar with the headword (Brezina & Gablasova, 2015, p. 4). A further issue with the principle of word families is the assumption that learners possess adequate morphological skills, which, some argue, is arrived at much later than the knowledge of inflectional word relationships (Gardner & Davies, 2014).

Another frequently cited limitation is based on Engels's (1968) study exploring the range (i.e., occurrence across texts) and coverage (i.e., proportion of a text covered by a certain number of words) of the GSL. In his study, Engels (1968) analysed 10 randomly selected texts of 1,000 words each and reported that the first 1,000 word families appeared in up to 9 out of 10 texts and accounted for approximately 70% of the words in the texts, while the second 1,000 word families

failed to occur across enough texts and represented less than 10% of the words. Engels (1968) thus criticised the second 1,000 word families for their low coverage and a limited range. His findings are, however, challenged by Gilner (2011), who points out that Engels's (1968) methodology is defective and hence his conclusion that the GSL lacks coverage and range is flawed. Specifically, Gilner (2011) points out that even if the ten 1,000-word texts used in Engels's (1968) study contained exclusively GSL words uniformly distributed across the texts, each GSL item could not appear in more than three texts. Gilner (2011) thus claims that it is unrealistic to expect an informative measure of range as the texts used in Engels's (1968) study were not sufficiently long for the second 1,000 word families to occur, and concludes that Engels's (1968) methodology "is clearly defective and the paper's conclusion that the GSL lacks range is evidently coerced by its own methodology" (p. 71). Gilner (2011) has, therefore, shown that although the GSL has often been criticised on grounds of range and coverage following Engels's (1968) study, these criticisms are unjustified.

The methodological principles involving subjective measures have also attracted criticism. Among these is the ease of learning. According to this principle, several words were included based on the similarity of the word form without meeting the criterion of frequency, meaning that the GSL contains vocabulary items which may be easily learned by learners, but which are relatively infrequent in spoken or written contexts (Brezina & Gablasova, 2015). The criterion of necessity, which was intended to ensure that all necessary ideas are covered in the wordlist, can also be regarded as problematic. This is mainly due to the rapid changes in the modern society, meaning that it may be difficult to establish what necessary ideas are. For instance, according to West (1953), despite its relatively low frequency, the verb *preserve* (food) represented a necessary idea as it subsumes others (e.g., *freezing* and *canning*) which cannot be explained without the use of the hypernym *preserve*. However, a present-day learner may find the verb *preserve* less important (Brezina & Gablasova, 2015). Another subjective criterion used by West relates to stylistic and emotional neutrality intended to result in only stylistically unmarked words being selected for inclusion in the GSL, which serve the function of communicating ideas rather than expressing emotions. The issue with this criterion is two-fold: it presupposes that neutral expression of ideas is the primary language function sought by English learners and it excluded some stylistically marked vocabulary items with high frequency (Brezina & Gablasova, 2015). Due to the limitations of the GSL, there have been several attempts to replace this list, discussed next.

## **2.2. The New General Service List**

To mark the 60th anniversary of West's GSL, Browne et al. (2013) announced the creation of a New General Service List (NGSL). The aims of introducing the NGSL were to update and expand the size of the source corpus (compared to the GSL) intended to increase the validity of the word list; to compile a word list of the most important high-frequency words useful for second language learners of English, which would also give as high coverage of English texts as possible with as few words as possible; to create a word list based on a clearer definition

of what constitutes a word; and to generate a discussion among scholars and teachers so as to revise the list (Browne, 2014).

The NGSL is based on a modified lexeme approach. Traditionally, the term “lexeme” refers to “a group of related forms which share the same meaning and belong to the same word class (part of speech)” (Biber et al., 1999, p. 54); in other words, traditional lexemes include inflected forms of words but count homographs separately. For example, the polysemous noun *time* can refer to a period or frequency. This difference in meaning would be distinguished in the lexeme approach and *time* would count as two lexemes. Unlike the traditional definition of lexeme, Browne et al. (2013) counted the headword in all its various parts of speech and included all inflected forms. For instance, *list* would include the noun and verb *lists*, the verb and adjective *listed*, the noun and verb *listing*, and the noun *listings*. This approach differs from the word family approach taken in the GSL by not including any of derived forms of a headword.

The NGSL is derived from a 273-million-word sub-section of the 2-billion-word Cambridge English Corpus (CEC). Specifically, the following sub-corpora of the CEC were utilised: learner, fiction, journals, magazines, non-fiction, radio, spoken, documents and TV. A number of computational procedures were then used to combine the frequencies from the selected sub-corpora and to adjust for differences in their relative sizes. The final stage of the NGSL creation involved a series of meetings between Browne et al. and Paul Nation regarding any potential improvement of the list. Following this, the NGSL was compared to West’s GSL, the BNC and COCA to ensure that important words were either included or excluded as necessary (Browne, 2014). The resulting NGSL contains approximately 2,800 high-frequency words.

The NGSL has addressed some of the criticisms of the GSL, namely, its basis on a relatively small and dated corpus. However, the issue of different meanings caused by the modified lexeme approach collapsing various grammatical classes (and thereby words displaying different meanings) becomes apparent from the following vocabulary entries in the NGSL: *figure* (having a different meaning as a verb and noun), *mean* (having a different meaning as an adjective, noun and verb), *novel* (having a different meaning as an adjective and noun) or *relative* (having a different meaning as an adjective and noun). Hence, although the NGSL can be seen as an advance on the GSL in terms of the corpus size and currency, the above examples of the different meanings caused by the modified lexeme approach illustrate that this approach to a word list creation remains its potential limitation.

### **2.3. The New General Service List**

The New General Service List (new-GSL) developed by Brezina and Gablasova (2015) is another response to the issues identified with West’s (1953) GSL, particularly the criticisms concerning its age, coverage, utility, the qualitative criteria involved in the selection of the GSL items and its organising principle based on word families. To address these issues, the new-GSL was derived from four language corpora<sup>1</sup> totalling over 12 billion words reflecting current language. Furthermore, the vocabulary items contained in the new-GSL were extracted by

purely quantitative measures using lemmas. A lemma refers to all inflectional forms related to one stem which belong to the same part of speech (Kučera & Francis, 1967, p. 1). In other words, according to a lemma principle, words with the same base and different grammar (e.g., the singular and plural forms of the same noun or the present and past tense of the same verb) are grouped together. To illustrate the difference between the word family principle adopted in West's GSL and the lemma principle adopted for the compilation of the new-GSL, Brezina and Gablasova (2015, p. 4) use the example of the verb *develop*. A lemma with the headword *develop* includes the inflectional verb forms *develops*, *developed* and *developing*, whereas a word family with the same headword would additionally include the nominal forms *development*, *developments*, *developer* and *developers* as well as adjectival derivatives *undeveloped* and *underdeveloped*. As can be seen from this example, the verbal and adjectival uses of *developing* are subsumed in one word family. Hence, the use of lemmas was intended to result in a pedagogically useful word list aimed at beginner learners of English who may possess limited word building skills. The resulting new-GSL comprises 2,494 lemmas representing significantly fewer forms than West's GSL with a comparable text coverage of around 80% of the texts in the four source corpora (Brezina & Gablasova, 2015).

Therefore, it can be said that the new-GSL represents a more up-to-date picture of contemporary English than West's GSL. Also, its foundation on lemmas instead of word families means that the new-GSL may be pedagogically more useful to learners with a lower level of proficiency as it eliminates the issue of meaning caused by the fact that grammatical parts of speech are not considered by word families as well as the difficulties that derivational word forms in word families may pose for these learners. Nevertheless, a potential limitation of the new-GSL is the fact that predominantly the British variety of English is included, which does not reflect the international character of English as an international language. This has, however, been addressed by Brezina and Gablasova (2015, p. 19) who note that preliminary findings of their study replicating this research using American English corpora indicate that "there is surprisingly small variation between the two varieties in the most frequent vocabulary". A further criticism could be seen in the use of purely quantitative approach for the creation of the new-GSL; as a result, it "may not include words that are not very high in frequency in written language but seem to be useful for L2 learning purposes such as *hey*, *hi*, and *ok*" (Dang & Webb, 2016b, p. 136).

#### **2.4. BNC2000 and BNC/COCA**

The BNC2000 (Nation, 2006) contains the most frequent 2,000 word families from Nation's (2006) 14 BNC lists (each containing 1,000 word families). The 14 BNC word family lists were derived from the 100-million-word BNC comprising predominantly written sources, which formed 90% of the corpus. Their creation was guided by the criteria of frequency, range (i.e., occurrence across texts) and dispersion (i.e., evenness of distribution across different texts). These criteria were applied to a list of lemmas (not word families) obtained from the BNC, which were subsequently used to divide the data into the 14 BNC word family

lists, with some word families containing several lemmas. For example, the word family of *abbreviate* subsumes the following members: *abbreviate*, *abbreviates*, *abbreviating*, *abbreviated*, *abbreviation* and *abbreviations*, and consists of two lemmas (i.e., *abbreviate* with four members and *abbreviation* with two members) (Nation, 2006, p. 63). In addition, some subjective judgements were made. These involved the inclusion of common spoken words such as months or weekdays, for instance, despite their lower frequencies in the corpus. This was to minimise the bias resulting from the more formal written component of the BNC.

One of the limitations of the BNC2000 could be seen in its bias towards the British variety of English due to its derivation from the BNC. Furthermore, owing to the more formal nature of the written texts contained in the BNC, the BNC2000 may be less suitable in contexts with young learners aiming to improve their conversational skills. To address these shortcomings, the BNC2000 was later updated by including the most frequent word families from the COCA, leading to the BNC/COCA word lists.

The BNC/COCA word lists consist of 29 lists of word families: 25 of the lists contain 1,000 word families each (based on the criteria of frequency and range) and the remaining four lists contain a list of proper names, e.g. *Beatles*, *Kuwait* or *Melissa* (21,662 word families); a list of marginal words including swear words, exclamations and letters of the alphabet, e.g., *oh*, *ouch* or *grr* (38 word families); a list of compounds, e.g., *onlooker*, *postman* or *snowstorm* (3,108 word families); and a list of acronyms, e.g., *ETA*, *TBC* or *UFO* (1,083 word families). The BNC/COCA was created from a specially designed 10-million word corpus comprising six million words of spoken English and four million words of written English obtained from the BNC, the COCA and to a lesser extent from the Wellington Corpus of Spoken New Zealand English (WSC). The inclusion of the American and New Zealand varieties of English (derived from COCA and WSC, respectively) was intended to avoid bias towards British English. The spoken English component of the source corpus represented, for example, face-to-face and telephone conversations, movies and TV programmes, while the written English component was based on texts for young children and general fiction, including letters or school journals, for instance. The composition of the source corpus was intended to ensure that different varieties of English were represented and that the word list was suitable for L2 learners by reflecting these learners' needs for spoken language, informal language and language used by children. The above-described source corpus was used for the creation of the first 2,000 word families, whereas the third 1,000 word families were based on COCA/BNC rankings after removing the first 2,000 word families. Although the creation of the list was guided by range and frequency, lower-frequency words (e.g., numbers and months) were also included if they were considered useful for L2 learners.

One of the strengths of the BNC/COCA lies in “its teaching-oriented purpose and its derivation from a corpus with a balance between spoken and written texts from different sources, and different varieties of English” (Dang & Webb, 2016b, p. 135). However, its organising principle based on word families means it may be less useful for learners at a beginner level who possess limited morphological knowledge.

## 2.5. Essential Word List (EWL)

The Essential Word List (EWL) (Dang & Webb, 2016a) is intended to provide a starting point for English learners' lexical development at a beginner level. The creation of the EWL was based on the GSL, BNC2000, BNC/COCA and new-GSL "by including the best items in terms of lexical coverage" from these existing word lists (Dang & Webb, 2016a, p. 155). The rationale behind combining these four lists was to achieve a greater coverage than any one of these lists. In contrast to the other lists, the EWL is based on an expanded version of lemmas, so-called flemmas, referring to lemmas which unlike the traditional definition of lemmas do not separate parts of speech. For example, the verb *developing* and the adjective *developing* represent different lemmas due to their different parts of speech, but as a result of their identical forms are members of the same flemma. The flemmas were used as an organising principle of the four lists forming the master list used as a basis for the EWL creation, whereby the word families and lemmas in the four lists were converted into flemmas and repeated items were excluded.

The resulting EWL contains 800 items comprising 624 lexical and 176 function words, which ought to enable learners at a beginner level to recognise over 60% of English words commonly used in both written and spoken texts. The EWL is divided into 13 sub-lists (according to decreasing mean coverage), each containing 50 headwords (except sub-list 13 comprising 24 headwords). The size of the sub-lists is intended to reflect the needs of individual English language courses, meaning that the sub-lists should be "small enough to fit into individual courses within an English language program" (Dang & Webb, 2016a, p. 167), while the division into sub-lists aims to increase learning effectiveness by exposing learners to the most useful items first. Another strength of the EWL relates to the use of flemmas as an organising principle, which may be useful for learners with limited receptive knowledge of inflectional and derivational word forms due to grouping different parts of speech (McLean, 2018), meaning that the EWL's organising principle does not require the learners to possess sophisticated morphological knowledge. Furthermore, its division into lexical and function words and the manageable sizes of the sub-lists may be regarded as a potential strength of the EWL. These characteristics make the EWL particularly suitable for learners at a beginner level (Dang & Webb, 2016a).

## 3. Summary and conclusion

Despite their common aim of providing a compilation of the most frequently occurring vocabulary in everyday contexts, the above-reviewed general word lists differ in a number of ways, including their age, size, source corpora and unit of counting. A summary of the main features of the above-discussed lists is provided in Table 1.

This article has provided an up-to-date overview of general word lists currently available with the aim of assisting EFL/ESL practitioners in making decisions with regard to the choice and utility of these word lists in their practices. It has described the characteristics of the word lists and highlighted the differences between them. The various characteristics of the word lists ought to be considered

Table 1. Overview of General Word Lists

Word list	Published	Source	Criteria	Unit of counting	Size
<b>GSL</b>	1953	5-million-word corpus of written English	Frequency and subjective measures (ease or difficulty of learning; necessity; stylistic level; intensive and emotional words)	Word family	2,000 word families (two word lists each containing 1,000 word families)
<b>NGSL</b>	2013	273 million words of the 2 billion word Cambridge English Corpus	Carrolls' dispersion measure (D2); estimated frequency per million (Um); standard Frequency Index	Modified lexeme	Approximately 2,800 words
<b>new-GSL</b>	2015	12-billion words drawn from: the LOB corpus; the BNC; the BE06 Corpus; EnTenTen12	Frequency; dispersion; distribution	Lemma	2,494 lemmas
<b>BNC2000</b>	2006	100-million-word BNC (90% written texts, 10% spoken texts)	Frequency; range; dispersion; subjective judgements	Word family	2,000 word families
<b>BNC/COCA</b>	2012	10-million-word corpus (60% spoken English from TV programmes and movies; 40% written English from fiction and children's texts)	Range; frequency	Word family	29 lists 3,000 word families
<b>EWL</b>	2016	GSL BNC2000 BNC/COCA2000 new-GSL	Mean coverage; practicability; change in the lexical coverage curve; lexical coverage	Flemma	13 lists 800 headwords (624 lexical and 176 function words)

by practitioners in their pedagogical practices. In particular, the organising principle or unit of counting is an important criterion as it “needs to represent the kind of word knowledge needed by the end-users of the list” (Nation, 2016, p. 8). The GSL, BNC2000 and BNC/COCA used word families as an organising principle. The new-GSL is based on lemmas, while the NGSL utilised modified lexemes and the EWL used flemmas as a unit of counting. Each of these principles has certain advantages as well as drawbacks and is suitable for learners at different levels of proficiency reflecting their morphological awareness. Specifically, word lists drawing on the word family principle may be more suitable for higher proficiency students, while those using the lemma, modified lexeme and flemma approach are likely to be more helpful for lower level learners. Furthermore, the size of the list as well as the inclusion of sub-lists should be manageable for the learners and should thereby also reflect the learners’ linguistic level and needs. In sum, practitioners should always consider the various characteristics of the word lists in relation to their group of learners and the choice of a word list should be determined by the learners’ level and linguistics needs.

## Note

1. These are as follows: the Lancaster-Oslo-Bergen Corpus (LOB), the British National Corpus (BNC), the BE06 Corpus of British English (BE06) and the English Web Corpus EnTenTen12.

## References

- Anthony, L. (2013). *AntWordProfiler*. [Computer Software]. Retrieved from <http://www.laurenceanthony.net/software>
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22.
- Browne, C. (2014). The New General Service List Version 1.01: Getting Better All the Time. *Korea TESOL Journal*, 11(1), 35–50.
- Browne, C., Culligan, B., & Phillips, J. (2013). New General Service List. Retrieved from <http://www.newgeneralservicelist.org/>
- Cobb, T. (n.d.). *VocabProfiler*. [Computer Software]. Retrieved from <https://www.lectutor.ca/vp/eng/>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213–238.
- Dang, T. N. Y., & Webb, S. (2016a). Making an essential word list for beginners. In *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.

- Dang, T. N. Y., & Webb, S. (2016b). Evaluating lists of high-frequency words. *International Journal of Applied Linguistics*, 167(2), 132–158.
- Engels, L. K. (1968). The fallacy of word-counts. *International Review of Applied Linguistics in Language Teaching*, 6(3), 213–231.
- Faucett, L., Palmer, H., Thorndike, E. L., & West, M. (1936). *Interim report on vocabulary selection*. London: P.S.King and Son, Ltd.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, 23(1), 65–83.
- Kučera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Lessard-Clouston, M. (2013). Word lists for vocabulary learning and teaching. *The CATESOL Journal*, 24(1), 287–304.
- McLean, S. (2018). Evidence for the Adoption of the Flemma as an Appropriate Word Counting Unit. *Applied Linguistics*, 39(6), 823–845.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How Large a Vocabulary Is Needed For Reading and Listening? *The Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from [https://www.victoria.ac.nz/\\_\\_data/assets/pdf\\_file/0004/1689349/Information-on-the-BNC\\_COCA-word-family-lists-20180705.pdf](https://www.victoria.ac.nz/__data/assets/pdf_file/0004/1689349/Information-on-the-BNC_COCA-word-family-lists-20180705.pdf)
- Nation, I. S. P. (2016). *Making and Using Word Lists for Language Learning and Testing*. Amsterdam/Philadelphia: John Benjamins.
- Nation, I. S. P., & Heatley, A. (1994). *Range: A program for the analysis of vocabulary in texts*. [Computer Software]. Retrieved from <https://www.wgtn.ac.nz/lals/about/staff/paul-nation>
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London: Continuum International Publishing Group.
- Richards, J. C. (1974). Word Lists: Problems and Prospects. *RELC*, 2, 69–84.
- Stoeckel, T. (2019). An Examination of the New General Service List. *Vocabulary Learning and Instruction*, 8(1), 53–61.
- West, M. (1953). *A general service list of English words*. London: Longman.