Article

Evaluating lists of high-frequency words: Teachers' and learners' perspectives

Thi Ngoc Yen Dang [ID]
University of Leeds, UK

Stuart Webb
University of Western Ontario, Canada

Averil Coxhead [ID]
Victoria University of Wellington, New Zealand

Abstract

With a number of word lists available for teachers to choose from, teachers and students need to know which list provides the best return for learning? Four well-established lists were compared and it was found that BNC/COCA2000 (British National Corpus / Corpus of Contemporary American English 2000) and the New General Service List (New-GSL) provided the greatest lexical coverage in spoken and written corpora. The present study further compared these two lists using teacher perceptions of word usefulness and learner vocabulary knowledge as the criteria. First, 78 experienced teachers of English as a second language / English as a foreign language (ESL/EFL) rated the usefulness of 973 non-overlapping items between the two lists for their learners. Second, 135 Vietnamese EFL learners completed 15 yes/no tests which measured their knowledge of the same 973 words. Teachers perceived that the BNC/COCA2000 had more useful words. Items in this list were also better known by the learners. This suggests that the BNC/COCA2000 is the more useful high-frequency wordlist for second language (L2) learners.

Corresponding author:
Thi Ngoc Yen Dang, School of Education, University of Leeds, Hillary Place, Woodhouse Lane, Leeds, LS2 9JT, UK.
Email: T.N.Y.Dang@leeds.ac.uk

I Introduction

Second language (L2) learners, especially those studying in English as a foreign language (EFL), have less exposure to the target language and less learning time than children learning their first language (Muñoz, 2008; Webb & Nation, 2017). Identifying which words that L2 learners should learn first is particularly important, because it helps them to get the best return for their learning effort (Nation, 2013). There are a small number of high-frequency words (around 2,000 items) (e.g. *think, alright, important*) that cover from 70% to 90% of the words in different kinds of texts (e.g. newspapers, general conversation, TV programs, and academic texts) (Coxhead, 2000; Dang & Webb, 2014; Nation, 2004). Knowledge of high-frequency words is important because it may allow learners to recognize a large proportion of words in different spoken and written texts. Such knowledge provides a solid foundation for learners to acquire words at lower frequency levels and achieve a high and stable degree of comprehension. For this reason, high-frequency words have been widely accepted as the crucial starting point for L2 vocabulary learning (Nation, 2013; Schmitt, 2010). Several corpus-based lists of high-frequency words have been developed such as General Service List (West, 1953), CELEX lists (Dutch Centre of Lexical Information; Baayen, Piepenbrock, & van Rijn, 1995), BNC2000 (British National Corpus 2000; Nation, 2006), COCA lists (Corpus of Contemporary American English list; Davies & Gardner, 2010), BNC/COCA2000 (Nation, 2012), SUBTLEX lists (Subtitles-based word frequencies; Brysbaert & New, 2009; van Heuven et al., 2014), New General Service List (Browne, 2014), and New General Service List (Brezina & Gablasova, 2015). These lists are useful for setting learning goals, designing learning materials and activities, and developing tests (Nation, 2016). To the best of our knowledge, little research has been conducted to examine the effectiveness of implementing these corpus-based high-frequency word lists in language classrooms. However, empirical studies with Coxhead's (2000) Academic Word List (AWL) have indicated that this list is a useful tool for learners (e.g. Banister, 2016; Lesaux et al., 2010; Townsend & Collins, 2009). As the AWL is a corpus-based word list, these findings highlight the value of corpus-based word lists for language learning and teaching.

Given that there are several different high-frequency word lists, one question that arises is which list is the most useful for L2 learners. Five studies have been conducted to address this question (Brezina & Gablasova, 2015, Browne, 2014; Dang & Webb, 2016; Gilner & Morales, 2008; Nation, 2004). Each study used lexical coverage as the sole criterion for determining which list is best. Lexical coverage refers to the percentage of words covered by items from a particular word list in a corpus (Nation & Waring, 1997). Dang and Webb's (2016) study was the most comprehensive because it compared the coverage of a larger number of word lists in a larger number of corpora, and these corpora had a great degree of diversity in types of texts, sizes, and varieties of English. Their comparison showed that Nation's (2012) BNC/COCA2000 accounted for the largest coverage while Brezina and Gablasova's (2015) New-GSL included the largest number of frequent items. This suggested that if one of the lists was used as a whole, the BNC/COCA2000 may provide the greatest value for learners; however, if only a proportion of the list was used, the New-GSL might have the greatest value.

Although lexical coverage is an important criterion to evaluate corpus-based word lists, to make these lists more relevant to L2 learning and teaching, list evaluation should involve their end-users— learners and teachers. Unfortunately, no studies have involved these agents in the evaluation of high-frequency word lists. To address this gap, the present study used learner vocabulary knowledge and teacher perceptions of word usefulness to further compare the BNC/COCA2000 and the New-GSL. The findings should indicate which high-frequency word list is more useful for L2 learners.

1 Background

Several high-frequency word lists have been developed for L2 learners. West's (1953) General Service List (GSL) is the oldest and most influential list. The GSL words were selected from a five million word corpus of written texts based on six criteria (frequency, ease of learning, necessity, coverage, stylistic level, and emotional neutrality). Research has shown that the GSL words account for around 70–90% of the words in different kinds of text such as academic writing (Coxhead, 2000), academic speech (Dang & Webb, 2014), movies (Webb & Rodgers, 2009), and novels (Nation, 2006). With its impressive coverage, the GSL has had a great impact on L2 vocabulary learning, teaching, and research (Webb & Nation, 2017). However, the GSL is not without limitations. First, this list was derived from a corpus which was made up of texts from the 1930s and thus may not fully reflect current vocabulary (Carter & McCarthy, 1988). Second, the GSL may be biased towards written English because it was developed solely from a written corpus (Carter & McCarthy, 1988). Third, the frequency and range of words beyond the first 1,000 words were not high enough to be included in a general high-frequency word list (Engels, 1968). Given these limitations, in recent years four word lists were created to improve on the GSL: Nation's (2006) BNC2000, Nation's (2012) BNC/COCA2000, Browne's (2013) New General Service List (NGSL), and Brezina and Gablasova's (2015) New-General Service List (new-GSL). The key features of these lists are presented in Table 1.

All of these lists consist of around 2,000 items. Lemmas were the unit of counting in Brezina and Gablasova (2015), flemmas were the unit of counting of Browne's (2013) NGSL, and word families were the unit of counting in Nation's (2006) BNC2000 and Nation's (2012) BNC/COCA2000. A lemma is a set of word forms which have the same stem and part of speech, but are different in inflections and/or spelling (Francis & Kučera, 1982). In other word, a lemma (*respond*) consists of a headword (*respond*) together with its inflected forms (*responds, responding, responded*). All members of a lemma belong to the same word class. Flemmas are similar to lemmas but do not take part of speech into account (Pinchbeck, 2014). For example, *smile* (v) and *smile* (n) are counted as two lemmas but one flemma. A word family (*respond*) includes a head word (*respond*), its inflected forms (*responds, responding, responded*) and closely related derivations (*respondent, respondents, responder, responders*). Similar to flemmas, word families do not distinguish between word classes.

Except for the BNC/COCA2000, all of these lists were derived from corpora consisting of mainly written texts. Brezina and Gablasova's (2015) New General Service List was developed from a purely quantitative approach; that is, using the average reduced frequency (Hlaváčová, 2006; Savický & Hlaváčová, 2002), which takes into account

Table 1. Key features of Nation's (2006) BNC2000, Nation's (2012) BNC/COCA2000, Browne's (2013) NGSL, and Brezina and Gablasova's (2015) New GSL.

Wordlists	Number of items[a]				Corpora	Selection criteria
	Word types	Lemmas	Flemmas	Word families		
West's (1953) GSL	13,451	n/a[b]	n/a[c]	2,168	5 million, 100% written	frequency, ease of learning, necessity, cover, stylistic level, and emotional neutrality
Nation's (2006) BNC2000	13,197	n/a[b]	n/a[c]	1,996	100-million, 90% written, 10% spoken	frequency, range, dispersion, subjective judgment
Nation's (2012) BNC/COCA2000	13,199	n/a[b]	n/a[c]	2,000	10-million, 40% written, 60% spoken	frequency, range, dispersion, subjective judgment
Browne's (2014) NGSL	8,205	n/a[b]	2,818	n/a[d]	274-million, 75.03% written, 24.97% spoken	frequency, dispersion, subjective judgment
Brezina and Gablasova's (2015) New General Service List	4,849	2,228	n/a[c]	n/a[d]	12-billion, 97.5% written, 2.5% spoken	frequency, dispersion, and distribution across language corpora.

Notes. [a]There is inconsistency across studies when reporting the number of word types from the same word list. It is because different authors may have slightly different views toward which types are considered as members of a word family/lemma. To achieve consistency in the comparison, in the present study, we considered word types sharing the same forms but different word classes (*smile* (v) and *smile* (n)) belonging to the same lemma/word family. Also, the headwords and members in the word lists were checked for consistency. [b]The number of lemmas is not reported because these lists are either word family or flemma list. [c]The number of flemmas is not reported here because these lists are either word family or lemma lists. [d]The number of word families is not reported because these lists are either lemma or flemma lists.

both the absolute frequency of a word and its distribution in the corpus, as the selection criterion. However, the development of the other lists also included some subjective criteria for word selection apart from these objective criteria. Common spoken words (e.g. *goodbye, ok, oh*), weekdays, months, numbers, letters, and names of countries were included in the Nation's (2006) BNC2000 and Nation's (2012) BNC/COCA2000 despite not meeting the frequency, range, and dispersion—Juilland's D (Juilland & Chang-Rodríguez, 1964)—criteria. This was to ensure that Nation's lists were appropriate for L2 learning and teaching. Browne (2013) claimed that feedback from teachers and learners was sought to perfect his list, but no information about the procedure of getting the feedback was provided.

2 Which high-frequency word list provides the greatest lexical coverage?

To our knowledge, five studies have explicitly compared the GSL with more current high-frequency word lists. Nation (2004) and Gilner and Morales (2008) reported that

the GSL did not provide as much coverage as the BNC2000. In contrast, Browne (2014) found that the GSL provided higher coverage than Browne's (2014) NGSL and Brezina and Gablasova's (2015) New-GSL in his fiction corpus, but lower coverage in his two magazine corpora. Similarly, Brezina and Gablasova (2015) found that the GSL covered a larger number of words in the Lancaster-Oslo-Bergen Corpus, British National Corpus, and the BE06 Corpus of British English than the New-GSL but lower coverage in the EnTenTen corpus.

Nation (2004), Gilner and Morales (2008), Browne (2014), and Brezina and Gablasova (2015) only compared three or fewer high-frequency word lists in no more than four corpora, and most of the corpora used for validation of their studies contained only written texts. To address these limitations, Dang and Webb (2016) compared the coverage of all four high-frequency word lists—the GSL, the BNC2000, the BNC/COCA2000, and the New-GSL— in 18 corpora. These corpora represented a wide range of spoken and written discourse types and 10 different varieties of English. Their results showed that the BNC/COCA2000 provided the highest coverage, but that the core of the New-GSL provided higher coverage (when an equal number of words from each list were compared, the New-GSL provided greater coverage). Overall, Dang and Webb' study indicated that the BNC/COCA2000 and the New-GSL provided the greatest lexical coverage and therefore might be the most useful lists for L2 learners.

Lexical coverage is an important criterion to evaluate high-frequency word lists because it is closely related to comprehension; the more words that are known in a text, the more likely that someone will understand the text (Schmitt, Jiang, & Grabe, 2011). In other words, the greater coverage that a word list provides, the more likely that list will help learners to comprehend spoken and written discourse.

It is important to note that using lexical coverage as the criteria for evaluating word lists has advantages and disadvantages. The advantage is that the lexical coverage of a word list in corpora indicates the value of the word list if learners were to encounter all the words in that corpus. If the corpora have a great deal of overlap with the language that students encounter, then the lexical coverage of the word list in the corpora will clearly indicate the value of the words. However, there is likely to be a lot of variation between the language that makes up a corpus, and the language encountered by learners in different contexts. Thus, lexical coverage may provide an indication of the usefulness of a word list. However, of the extent to which the words that make up a word list are relevant to learners in a *particular* context will likely vary (Milton, 2009). For example, Stein's (2017) points out that some items in Brezina and Gablasova's (2015) New-GSL may not be relevant to EFL beginners, and therefore, raises the concern that teachers and learners may not see clearly the contribution of corpus-based word lists to their teaching and learning. Stein's (2017) concern is supported by the findings of subsequent studies. Dang, Webb, and Coxhead, under review) examined the relationship between lexical coverage, learner knowledge, and teacher perceptions of the usefulness of high-frequency words. They found that although lexical coverage significantly correlated with the other two factors, the correlations were small: $r = .20$ (learner knowledge) and $r = .23$ (teacher perception of word usefulness). Similarly, He and Godfroid (2019) found a moderate correlation between the frequency of academic words in the COCA and COCA-Academic corpus and teacher perceptions of the usefulness of these words ($r = .44$).

These results indicated that while lexical coverage from corpora is a key criterion to evaluate corpus-based lists of high-frequency words, list evaluation should involve their end-users—learners and teachers.

3 Learner vocabulary knowledge

Language is a complex system in which different elements are intertwined with each other, but language also has patterns (Beckner et al., 2009). With regard to vocabulary, studies measuring the vocabulary knowledge of L2 learners in different contexts (e.g. Henriksen & Danelund, 2015; Laufer, 1998; Matthews & Cheng, 2015; Nguyen & Webb, 2017; Stæhr, 2008; Webb & Chang, 2012) showed that learners knew more high-frequency words than those at lower frequency levels. Experimental studies (e.g. Ellis, 2002; Ellis, Simpson-Vlach, & Maynard, 2008; Hernández, Costa, & Arnon, 2016) also indicated that L2 learners are sensitive to word frequency. This suggests that high-frequency words are likely to be learned before those at lower frequency levels. Moreover, as L2 construction is influenced by various factors (e.g. cognition, consciousness, experience, embodiment, brain, self, human interaction, society, culture and history) (Beckner et al., 2009), measuring the vocabulary knowledge of L2 learners may reveal individual experience and the extent to which students are exposed to the target language in a specific context (Schmitt, 2010). Taken together, previous research on learner knowledge provides an indication of the relevance and value of words to learners in particular contexts. This suggests that learner knowledge would also be a useful criterion along with lexical coverage in the assessment of high-frequency words.

4 Teacher perceptions of word usefulness

Teacher perception of word usefulness is another important criterion for evaluating the word lists. Words selected for learning should be as useful as possible so that the learning time is well spent (Laufer & Nation, 2012). That is, words that are useful for learning should be the words that are encountered frequently in speech and written text and so aid comprehension while also having value in helping students to communicate effectively in speech and writing. While words with high lexical coverage of corpora are likely to be useful for learners, other situational factors (e.g. learning purposes, tests, curricular, materials, parents/society expectations, and students' characteristics) also play a role in determining the usefulness of words for L2 learners (Gerami & Noordin, 2013; Lau & Rao, 2013; Zhang, 2008). As these factors are intertwined, teacher perceptions of word usefulness can provide an implicit indication of the influence of different factors on the value of a word for learning. Teachers play a significant role in L2 vocabulary learning, especially in EFL contexts (Dang et al., under review); Laufer, 2003; Schmitt, 2008). Their direct involvement in the teaching and learning process may allow teachers to have a strong understanding of which words are needed for communication in that context. Thus, teacher perceptions of word usefulness can provide useful insight into the value of the items that make up a word list. Research with other languages (French, Italian, and Turkish) has shown that teacher perceptions (Tidball & Treffers-Daller, 2008) or lexical coverage plus teacher perceptions (Bardel, Gudmundson, & Lindqvist, 2012; Tidball &

Treffers-Daller, 2008) is better at determining the lexical sophistication in speech produced by L2 learners than the lexical coverage of the words in corpora alone. Teacher perception has been used in the development and validation of academic vocabulary lists (He & Godfroid, 2010; Simpson-Vlach & Ellis, 2010), but no studies have used teacher perceptions to validate high-frequency word lists. In fact, no research has used both learner vocabulary knowledge and teacher perceptions of word usefulness in the validation of high-frequency word lists. As a result, researchers (Dang, 2020; Gilner, 2011; Nation, 2016) have called for using other criteria to supplement lexical coverage in word list validation to move the field forward.

II The present study

The present study is the first attempt to use information from teachers and learners to supplement corpus-based information in the evaluation of high-frequency word lists. Expanding on Dang and Webb's (2016) study, the present study further compared Nation's (2012) BNC/COCA2000 and Brezina and Gablasova's (2015) New-GSL by using two criteria: (a) L2 learner vocabulary knowledge and (b) teacher perceptions of the usefulness of the words for basic functions in English. The research involved the participation of 135 L2 learners and 78 experienced teachers of English as a second language / English as a foreign language (ESL/EFL). The learners were from different proficiency levels and the instructors had experienced teaching in a wide range of EFL/ESL contexts. Therefore, this study is expected to provide an assessment of corpus-based lists of high-frequency words from the perspectives of learners and teachers. The study should shed light on the value of including the perspectives of teachers and learners in corpus-based word list validation, and thus, bring corpus-linguistics research together with other research strands—Second Language Acquisition and teacher cognition. Moreover, the findings should indicate which high-frequency word list is the most suitable for L2 learners. This in turn should help teachers and materials writers to select words for materials, activities, and tests for L2 learners.

The following research questions are addressed:

1. Which words do experienced English language teachers perceive as being most useful, words unique to Nation's (2012) BNC/COCA2000 or those unique to Brezina and Gablasova's (2015) New-GSL?
2. Which list accounts for a larger proportion of words known by L2 learners?

III Methodology

1 Participants

a Teachers. Seventy-eight English language teachers participated in this study: 25 EFL/ESL teachers who were native speakers of English, 26 Vietnamese EFL teachers, and 27 EFL teachers from varying countries. The native speakers of English had taught L2 learners from a wide range of first language (L1) backgrounds in ESL/EFL contexts (e.g. Arabic, Chinese, Ethiopian, French). The Vietnamese EFL teachers had experienced

Table 2. Nationalities and years of teaching experience of the teacher participants (n = 78).

ESL/EFL teachers who were native speakers of English	(n = 25)	Vietnamese EFL teachers	(n = 26)	EFL teachers from varying countries	(n = 27)
New-Zealander	13	Vietnamese	26	Indonesian	6
American	4			Malaysian	6
British	3			Iranian	2
Canadian	3			Japanese	2
Australian	2			Taiwanese	2
				Thai	2
				Chinese	1
				Greek	1
				Jordanian	1
				Kenyan	1
				Laotian	1
				Sri Lankan	1
				Venezuelan	1
Years of teaching experience	2–40 years (M = 13.12, SD = 9.35)	Years of teaching experience	2–22 years (M = 6.88, SD = 5.29)	Years of teaching experience	2–20 years (M = 8.63, SD = 4.64)

teaching English to Vietnamese EFL learners in Vietnam. The EFL teachers from vary-ing countries (e.g. Thai EFL teachers) had experience teaching English as a foreign lan-guage to learners who shared the same L1 background as them (e.g. Thai EFL learners) in their home countries (e.g. Thailand). All of the teacher participants had experience teaching English to L2 learners from beginner to advanced levels. The nationalities and years of teaching experience of these teachers are presented in Table 2. Given the diver-sity in these teachers' L1 backgrounds, teaching contexts, and experience, it was expected that they would provide a comprehensive assessment of the two high-frequency word lists.

b Learners. The learner participants were 135[1] Vietnamese EFL undergraduate students from 21 intact classes at six universities in Vietnam. They were enrolled in a range of academic majors (Table 3) and their years of studying English ranged from 2 to 15 years (M = 9.12; SD = 2.47). The learners were divided into three groups (pre-intermediate, intermediate, advanced) based on their scores in Schmitt, Schmitt, and Clapham's (2001) Vocabulary Levels Test (Table 4). Undergraduate students rather than postgraduate stu-dents were selected as the participants because high-frequency words are more relevant to the former group than the latter. According to Dunlea et al. (2018), to meet the gradu-ation requirements by the Ministry of Education, undergraduate students need to achieve at least the B1 level in the Common European Framework of Reference for Languages. To be admitted to postgraduate programmes, students need to obtain an undergraduate degree. Therefore, it is reasonable to expect that postgraduate students in Vietnam have already mastered at least the B1 level. Given this context, learning high-frequency words is more relevant to undergraduate students than postgraduate students.

Table 3. Learner's academic majors.

Academic majors	*n*
TESOL	86
Computer Sciences & Technology	31
Natural Sciences	13
Economics & Business	2
Law	2
Social Sciences & Humanities	1
Total	135

Table 4. Groups of learners (*n* = 135).

Group of learners (vocabulary level)	Number of learners	VLT score
Pre-intermediate	37	Scored from 50–80% at the 2,000-word level
Intermediate	50	Mastered the 2,000-word level
Advanced	48	Mastered at least the 3,000-word level

2 Target words

The target words were 973 non-overlapping headwords between the BNC/COCA2000 (545 headwords) and the New-GSL (428 headwords) (for information about these words, see Appendices 1–3 in the Supplementary data). That is, all words that were unique to a list were included as target items, while those that were found in both lists were not included as target items. This is because the current study aimed to compare the BNC/COCA2000 and the New-GSL. Comparing items appearing in both lists is then unnecessary because these items tell us about the similarities rather than the difference between the two lists. Moreover, words appearing in both lists are likely to be strong items and should be included in the list of high-frequency words for L2 learners; in contrast, words that are unique to each list tend to be not quite as strong as overlapping items because they may be the result of corpus differences (Nation, 2016; Nation & Hwang, 1995). Therefore, the non-overlapping items need further validation from learners and teachers.

Headwords were chosen as the unit of counting for the target words because headwords are usually the most frequent members of word families, and thus the most likely members to be known. This is supported by a corpus-based study which indicated that the headword was the most frequent member of 82% of the most frequent 1,000 word families in Nation's (2006) British National Corpus word lists (Brown, 2018). Using headwords also reflects the nature of L2 teaching and learning (Brown, 2018; Dang & Webb, 2016). That is, L2 teachers and learners usually receive lists of headwords without their inflections and derivations, and, therefore, are most likely to choose headwords to teach and learn first. Using headwords also helps to deal with the inconsistency in the number of word types reported by different studies (for more details, see Dang & Webb,

2016) and 'ensure a degree of coherence of organization and selection' (Gilner, 2011, p.68). The New-GSL lemma headwords were converted into the word family headwords by grouping lemma headwords belonging to the same word family together. For example, two lemma headwords *able* and *ability* were listed under the same word family headword *able*. Lemmas which shared the same forms such as *smile* (v) and *smile* (n) were classified as belonging to the same word families. There are three reasons for converting New-GSL lemma headwords to word family headwords. First, the present study is a follow study of Dang and Webb (2016) which compared the lexical coverage of the GSL, BNC2000, BNC/COCA2000, and the New-GSL. As three out of the four lists used word family as the unit of counting (GSL, BNC2000, and BNC/COCA2000), word family was chosen as the unit of counting in the present study. Second, in a pilot study with three teachers of English (one English L1 teacher, one Vietnamese L1 teacher, and one Chinese L1 teacher) and three Vietnamese EFL learners (one pre-intermediate, one intermediate, and one advanced), we used the BNC/COCA2000 word family headwords and the New-GSL lemma headwords as the target words. Feedback from the participants revealed that they were confused when rating lemmas from the same word family such as *achieve–achievement, construct–construction, demonstrate–demonstration*, and *effective–effectively* because they thought that they had to rate the same items repeatedly. Third, converting the New-GSL lemma headwords to word family headwords reduced the total number of target words and made it more feasible to recruit a larger number of participants for the present study.

a Teacher Likert surveys of target words. Ten surveys were developed to examine the teachers' perception of the usefulness of the 973 target words. In these surveys, the teacher participants would indicate in a five-point Likert scale the usefulness of each word in helping their students to perform basic functions in English. Point 1 on the scale was labelled as the least useful, and Point 5 the most useful. Seven of the surveys contained 97 target words and three surveys contained 98 target words. Stratified randomization was used to ensure that each survey had an equal proportion of BNC/COCA2000 words and New-GSL words. A sample of the surveys is presented in Figure 1.

The 10 surveys were in the Excel format and were emailed to each teacher. That is, each teacher would rate all 973 target words (545 headwords that are unique to the BNC/COCA2000 and 428 headwords that are unique to the New-GSL). This method of data collection allowed researchers to collect data from teachers in a wide range of geographic locations while causing minimal intrusion into their busy working life (Dörnyei & Taguchi, 2010). It also allows researchers to achieve a high rate of responses with valid data. If all 973 target words had been included in one online survey, it may have either discouraged teachers from taking part in the present study or resulted in fatigue effects when completing the survey. Distributing these target words into 10 short surveys for teachers to complete when they had time solved these problems. Additionally, emailing the surveys to each individual teacher enabled the researchers to better manage the progress of each participant.

The data collection with the teachers had several stages. First, an official invitation was sent to ESL/EFL teachers through different channels such as teacher networks or face-to-face meetings. Second, the first author set up one-on-one meetings (either

In the following section, please give your answer to the following question:

"TO WHAT EXTENT IS THE WORD USEFUL FOR YOUR STUDENTS?
(FOR GENERAL PURPOSES, NOT ACADEMIC PURPOSES)

By typing a number from **1** to **5** in the **degree of usefulness** column
(**1** is the **LEAST useful** and **5** is the **MOST useful**).

For example, if you think that knowing the word "haste" is not useful at all, knowing the word "lorry" is
relatively useful, and knowing the word "go" is very useful, your response will be as follows.

Number	Word	Degree of usefulness
1	haste	1
2	lorry	4
3	go	5

Figure 1. Sample of the surveys.

face-to-face, Skype, or Facebook meetings) with teacher participants to provide them with detailed instruction of how to complete the surveys. To avoid biasing the participants towards a certain word list, the names of the word lists from which the target words were taken was not mentioned to the participants. Third, to minimize intrusion into the teachers' busy working schedules, the teachers were given the flexibility to choose how often and how many surveys would be sent to them each time. The teachers downloaded the surveys, completed them, and emailed them back when they finished. To minimize the impact of the variation in the way that teachers responded to the surveys on the results of the study, the teachers were asked to complete the surveys as soon as possible but not to try to finish them all at the same time. After that, the results were checked, and the teachers were asked to provide further information if necessary.

b Learner yes/no tests of the target words. Fifteen yes/no tests were created to measure the learners' receptive knowledge of form-and-meaning relationship of the target words (see Figure 2). Form-and-meaning relationship was chosen because it is the most important aspect of vocabulary knowledge and acts as the foundation for further development of other aspects of vocabulary knowledge (Nation, 2013; Schmitt, 2010). The yes/no test format was chosen because it is the most suitable format to measure a large number of target words with a large number of participants in a limited period of time, which allows a high sampling rate for reliable estimation (Meara & Buxton, 1987, Read, 2000; Schmitt et al., 2011). A total of 480 pseudowords were included in the yes/no tests to minimize learners' overestimation of their vocabulary knowledge. Pseudowords (e.g. *freath*) are similar to real words in the language being tested (Meara & Buxton, 1987). They have been widely used as a means to deal with learners' overestimation of their vocabulary knowledge in the yes/no test format (Read, 2000). The use of pseudowords is based on

---

Original version

Với mỗi từ dưới đây, khoanh tròn: "Yes" nếu bạn biết ít nhất 1 nghĩa của từ
                                              "No" nếu bạn không chắc từ đó nghĩa là gì
Lưu ý, một số từ trong bài test <u>không tồn tại</u> trong tiếng Anh. Vì vậy xin bạn hãy thành thật khi đưa ra câu trả lời.

Ví dụ, nếu bạn biết ít nhất 1 nghĩa của từ "go", nhưng không biết nghĩa của từ "hapgood", câu trả lời của bạn sẽ như sau:
1. go   Yes   No   2. hapgood   Yes   No

English translation

For each word, circle "Yes" if you know at least one meaning of the word
                                    "No" if you do not know its meaning
Notes: some words in the test are <u>not real words</u> in English. Please be honest when giving the answers.

For example, if you know at least one meaning of the word "go", but you do not know the meaning of "hapgood", your answer will be as follows:
1. go   Yes   No   2. hapgood   Yes   No

Figure 2. Sample of the yes/no tests.

the assumption that if test-takers know all the words, they will tick 'Yes' to all the real words but 'No to all the pseudowords; if they tick 'Yes' to pseudowords, their overall test scores will be adjusted accordingly (Meara & Buxton, 1987) or data of participants who checked more than the acceptable percentage of pseudowords used in the tests will be removed (Schmitt et al., 2011).

The 973 target words and 480 pseudowords were distributed in 21 in-class tests in six universities in Vietnam. Thirteen of these tests had 97 items, and two contained 96 items. Stratified randomization was used so that each survey has around 36–37 BNC/COCA2000 words, 28–29 New-GSL words, and 32 pseudowords (for the number of BNC/COCA words, New-GSL, and pseudowords in each test, see Appendix 4 in the Supplementary data).

Permission was sought from the participants for this study. The tests had a paper-and-pencil format, which allowed one of the researchers to supervise this part of data collection, which increased the chances that the participants completed the tests and took the tests seriously. It also provided opportunities to meet the participants face-to-face and have follow-up participant checking about the options that they did not answer in the tests right after each test session. All instructions were in Vietnamese so that learners were clear about how to complete the tests.

One criticism of using the yes/no format is the face validity; that is, it does not require test takers to actually demonstrate their vocabulary knowledge and may then lead to the risk of test-takers not taking the test seriously (Nation & Webb, 2011; Read, 2000). However, previous studies have reported strong correlations between the vocabulary tests using yes/no test format with vocabulary tests using the multiple choice format (e.g. $r = .84$ (Anderson & Freebody, 1983); $r = .703$ (Meara & Buxton, 1987)) and the matching format (e.g. $r = .85$ to .88 (Mochida & Harrington, 2006)). This means students who got high scores in vocabulary tests using yes/no formats tend to get high scores in vocabulary tests using other formats. Additionally, Laufer (1992) found that the

correlation between reading comprehension and vocabulary knowledge measured by a vocabulary test in the yes/no test format ($r = .75$, $p < 0.0001$) is as strong as that measured by a vocabulary test in the matching format ($r = .5$, $p < 0.0001$). Schmitt et al. (2011) also found a moderate correlation between vocabulary knowledge measured by the yes/no test and scores on the reading comprehension test $r = .407$ ($p < .001$).

3 Analysing the teacher data

The teacher ratings in the survey data were analysed in two ways. The first analysis examined the usefulness of the 545 BNC/COCA words versus the 428 New-GSL words. The second analysis looked at the usefulness of the most frequent 428 of the 545 BNC/COCA words versus the 428 New-GSL words. The first analysis provides an assessment of the lists as a whole while the second analysis took the difference in the number of items in the BNC/COCA and New-GSL lists into consideration. This allowed us to determine the relative value of each item in the lists. To identify the most frequent 428 items of the 545 BNC/COCA2000 headwords, five steps were followed. First, the frequency of the 545 BNC/COCA headwords in each of the 18 corpora used in Dang and Webb (2016) was determined by running each corpus through RANGE with these headwords as the baseword list. Second, the coverage provided by each headword in each corpus was calculated by dividing the frequency of the headword by the number of running words in the corpus, and multiplying by 100. In the third step, the mean coverage provided by the headword in the 18 corpora was determined by adding the coverage provided by the headword in each corpus together, and then dividing by the number of corpora (18). Mean coverage was used to rank the headwords rather than the combined frequencies because combined frequencies would bias the results towards the findings of the largest corpora. In the fourth step, the 545 headwords were ranked according to their mean coverage in descending order. In the last step, the top 428 headwords were identified.

For each analysis, the same four steps were followed. First, items in the two sets of words used in the comparison were sorted in descending order by the mean score given by the teachers, and then by the standard deviation (*SD*) in ascending order. Second, the following indicators of usefulness were determined: (a) words with mean scores of 4 or above, (b) top 100 useful words, (c) top 200 useful words, (d) top 300 useful words, (e) top 400 useful words, and (f) top 500 useful words. Third, items that met these criteria were identified and selected for comparison. Finally, the proportions of the BNC/COCA2000 and the New-GSL words among the words that met each criterion were calculated and compared. A series of Z tests for the two population proportions were conducted to determine whether there existed a significant difference between these proportions.

4 Analysing the learner data

The learner test data were analysed by comparing the learners' knowledge of (a) the 545 BNC/COCA words versus the 428 New-GSL words and (b) the most frequent 428 items from each set of words. For each kind of analysis, the words known by 90% of the learners were identified. Then, the proportions of the BNC/COCA2000 and the New-GSL

words among these words were determined, and a series of Z tests for two population proportions were conducted to determine whether there existed a significant difference between the proportions of the two lists.

To ensure that the yes/no test results accurately estimated learners' vocabulary knowledge, following Schmitt et al.'s (2011) approach, only the data of 112 learners who ticked no more than 10% of the pseudowords were used for the analysis. Schmitt et al.'s approach was followed because the purpose of their study is similar to that of the present study, that is, to determine exactly which items from the word lists known by the participants who did not randomly guess or overestimate their vocabulary knowledge. In contrast, following other studies using correction formulas would not allow us to identify exactly which BNC/COCA words and New-GSL words were known by the learners because the formulas would only provide us with a figure estimating the overall number of target words known by the participants. For example, student S1 checked 624 words, including 15 pseudowords and 609 real words. Applying the correction formulas, the overall score (594) was calculated by subtracting the number of checked pseudowords (15) from the number of checked real words (609). Using correction formula would only reveal that this student knew a total of 594 out of 973 real words. It does not allow us to see exactly which BNC/COCA words and New-GSL words are counted as the 594 known items because the 609 real words checked by the participant had been adjusted.

The 10% cut-off point was chosen for three reasons. First, Schmitt et al. (2011) also adopted this cut-off point when using yes/no test format to measure L2 learners' knowledge of words from two different texts. Second, the 10% cut-off point is supported by the results of a preliminary analysis which compared the number of learners remaining when different percentage cut-off points of checked pseudowords were chosen (Table 5). Choosing stricter cut-off points (0%, 1%, or 5%) would result in a small number of learners either in total or in each group. In contrast, choosing a maximum of 10% error ensured that the present study had 112 learners in total with more than 30 learners in each group which makes it possible to apply statistical measures (Hatch & Lazaraton, 1991). This cut-off point also results in around 32 pre-intermediate learners and a good balance between the number of intermediate and advanced learners ($n = 40$).

Analysis showed that the 10% figure ensured that the yes/no test results were as reliable as the stricter cut-off points. Following Schmitt et al.'s (2011) approach, the current researchers conducted a series of independent-sample t-tests to compare the scores at the Vocabulary Levels Test (VLT) of 88 learners who checked no more than 5% of the pseudowords, with that of the 112 learners who checked no more than 10% of the pseudowords. The results showed no significant differences in the overall vocabulary levels score for the 5% set ($M = 55.65\%$, $SD = 19.11$), and the 10% set ($M = 54.73$, $SD = 18.38$), $t(198) = .344$, $p = 0.731$ (2-tailed). Normality was confirmed using Kolmogorov–Smirnov test for normality ($p > 0.05$), and visually assessed using Q-Q plots and boxplots. Similar analysis with the 2K, 3K, and 5K levels revealed the same results. These results suggested that choosing 5% or 10% did not make any difference in the VLT mean scores. Furthermore, a comparison between the VLT scores of the 112 learners who checked no more than 10% of the pseudowords with the total number of target words they indicated were known in the yes/no tests revealed that there was a linear relationship between the two variables. In particular, there were positive, strong, significant

Table 5. Number of learners at different cut-off points of checked pseudowords.

Checked pseudowords cut-off points	Number of learners in each group			
	Pre-intermediate	Intermediate	Advanced	Total
0%	0	0	0	0
1%	12	6	7	25
5%	25	28	35	88
10%	32	40	40	112
Original data	37	50	48	135

correlations between the learners' scores in the yes/no test and the VLT: $r = .85$ (2K level), $r = .78$ (3K level), $r = .69$ (5K), $r = .83$ (overall VLT score). Given the high validity of the VLT (Schmitt et al., 2001), the strong correlation between the learners' scores in the two tests suggested that the data of the 112 learners who ticked no more than 10% of pseudowords in the yes/no tests were accurate indicators of their vocabulary knowledge of the target words.

IV Results

The BNC/COCA2000 consistently made up a significantly larger proportion of words perceived as useful by teacher participants than the New-GSL. Table 6 presents the results of the analysis with all 78 teachers. When the teachers' ratings of 973 items (545 BNC/COCA2000 words and the 428 New-GSL words) were examined, depending on the criteria of usefulness, the BNC/COCA words accounted for 60.40% to 79.0% of the useful words while the New-GSL only accounted for 21% to 39.60%. Similarly, when the teachers' ratings of the most frequent 428 items from each set of words were compared, depending on the criteria of usefulness, the range of percentage of words from each list among the useful words rated by teachers was 61%—80% (BNC/COCA2000) and 20%—39% (New-GSL). The results of the Z-tests indicated that the differences were significant at $p < 0.05$ in all cases. Interestingly, as shown in columns 4 and 7, as the criterion of usefulness got stricter (from top 500 to top 100), the differences between the percentage of the BNC/COCA2000 and the New-GSL words among the most frequent words rated by the teachers become larger. Appendix 5 in the Supplementary data presents information about the percentage of the BNC/COCA2000 and New-GSL words in each 100-most useful word band rated by all teachers. In most cases, the BNC/COCA accounted for larger proportion of the useful words.

The same pattern was found with each group of teachers (EFL/ESL teachers who were native speakers of English, Vietnamese EFL teachers, and EFL teachers from varying countries). The BNC/COCA2000 always comprised a larger proportion of words rated as useful by the teachers (59–80%) than the New-GSL (20–41%) (for more details, see Appendices 6–8 in the Supplementary data). The results of the Z-tests revealed that except for the case of the top 100, 200, and 500 useful words rated by the EFL teachers from varying countries, the differences were always significant at $p < 0.05$. Similar to

Table 6. Percentage of the BNC/COCA2000 and New GSL among the words rated as useful words by all teachers (percentages).

	The whole sets of words			The most frequent 428 words from each set		
	BNC/COCA 2000	New-GSL	Difference in the percentage	BNC/COCA 2000	New-GSL	Difference in the percentage
Mean score of 4 or above	72.68	27.32	45.36	73.22	26.78	46.44
Top 100	79	21.0	58.0	80.0	20.0	60.0
Top 200	71.5	28.5	43.0	73.0	27.0	46.0
Top 300	66.0	34.0	32.0	67.0	33.0	34.0
Top 400	61.5	38.5	23.0	63.25	36.75	26.5
Top 500	60.4	39.6	20.8	61.0	39.0	22.0

the case of all teachers, for each group of teachers, as the criterion of usefulness got stricter, the differences between the proportion of the BNC/COCA2000 and the New-GSL among the most useful words rated by the teachers became larger (see Appendix 9 in the Supplementary data). It would be interesting to see which target words were consistently rated as the most useful by the teachers. Therefore, for each target word, we also counted the number of teachers indicating it as the most useful (having score of 5). Then, we ranked the 973 target words based on the number of teachers who indicated the word as most useful and identified the top 103[2] words. The results show 81.55% of these words (84 out of 103 words) are from the BNC/COCA2000 while only 18.45% (19 out of 103 words) are from the New-GSL. The results of the Z test indicated that the difference was significant at $p < 0.05$. Taken together, the results of the teacher surveys indicated that from the teachers' perspectives, the BNC/COCA2000 is a more useful high-frequency word list for L2 learners than the New-GSL. It is important to note that there were strong correlations between the ratings of each group of teacher: $r_{\text{English L1 teachers-Vietnamese L1 teachers}} = .81, r_{\text{English L1 teachers-Various L1 teachers}} = .83, r_{\text{Vietnamese L1-Various L1 teachers}} = .85, p < 0.001$. This indicates that the ratings of the teacher are very consistent.

The results from the learner data are less transparent than those from the teachers. Columns 2 and 3 of Table 7 present the results of the analysis of all 112 learners. No matter whether the full sets of words or only the most frequent 428 items from each set were compared, the BNC/COCA2000 always accounted for a larger percentage of known words (61.95%, 58.46%) than the New-GSL (38.05%, 41.54%). The differences were always significant at $p < 0.05$

Analysis with the data of intermediate and advanced learners revealed that there was no significant difference in the proportion of the BNC/COCA words and New-GSL words among the words known by these learners (see the last four columns of Table 7). This finding suggests that it is unclear which list is better known by intermediate and advanced learners. In contrast, as shown in columns 4 and 5 of Table 7, the BNC/COCA2000 always accounted for a significantly larger proportion of words known by

Table 7. Percentage of the BNC/COCA2000 and New-GSL words among the words known by the learner participants (percentages).

Sets of words used in the comparison	All learners (*n* = 112)		Pre-intermediate (*n* = 32)		Intermediate (*n* = 40)		Advanced (*n* = 40)	
	BNC/COCA 2000	New GSL	BNC/COCA 2000	New GSL	BNC/COCA 2000	New GSL	BNC/COCA 2000	New GSL
Full sets of words	61.95	38.05	67.36	32.64	56.40	43.60	52.94	47.06
most frequent 428 items from each set	58.46	41.54	65.44	34.56	52.82	47.18	51.27	51.27

pre-intermediate learners (67.36%, 65.44%) than the New-GSL (32.64%, 34.56%), and the differences were always significant at $p < 0.05$. As L2 learners tend to know more high frequency words than lower frequency words(e.g. Dang et al., under review) and tend to learn high frequency words first (Ellis, 2002), the results indicate that the BNC/COCA words seem to be learned before the New-GSL words.

When the learner data and teacher data were compared, there were 146 words known by at least 90% of the learners and indicated as being useful by the teachers (mean scores of 4 or above) (see Appendix 10 in the Supplementary data). Of these items, 108 words were from the BNC/COCA2000 and 38 words were from the New-GSL. The results of the Z test indicated that the difference was significant at $p < 0.05$. It is important to note that 7 out of 38 New-GSL words which are indicated as useful and known by most learners (*bathroom, bedroom, website, weekend, birthday, classroom, CD*) did not appear in the BNC/COCA2000 because the BNC/COCA lists include a separate list of transparent compounds and abbreviations.

V Discussion

Together, the data from teacher surveys and learner yes/no tests indicate that the BNC/COCA2000 is likely to be perceived as more useful by teachers and the items are likely to be learned earlier by L2 learners than the New-GSL. This suggests that the BNC/COCA2000 may be the more useful resource at least for EFL learners in Vietnam. There are two possible reasons for the superiority of the BNC/COCA over the New-GSL in terms of teacher perceptions and learner vocabulary knowledge. The first reason may be the result of the principles under which the two lists were developed. The New-GSL was created with a purely quantitative approach; that is, using the average reduced frequency (Hlaváčová, 2006; Savický & Hlaváčová, 2002), which takes into account both the absolute frequency of a word and its distribution in the corpus, as the selection criterion. In contrast, apart from these quantitative corpus-based criteria, the development of the BNC/COCA2000 also involved adding to the list the lexical items that did not meet these criteria but may be suitable for L2 learning and teaching purposes. A word list that is solely based on the information from corpora may miss items that have low frequency in corpora but are useful for L2 learning (Nation, 2016). For example, BNC/COCA words such as *alright, ok, exam, hello, goodbye, grade, pronounce, schedule, silence* were not

included in the New-GSL but were known by more than 90% of the learners and were rated as useful by teachers. The greater focus on L2 learning may explain why the BNC/COCA2000 had a larger number of words known by learners and perceived as being useful by teachers than the New-GSL. The second reason may be the corpora used to develop these lists. The BNC/COCA2000 was created from a corpus with a better balance of spoken texts (60%) and written texts (40%) and represents different varieties of English (British English, American English, and New Zealand English). In contrast, the New-GSL may be biased towards British, written English. Three out of the four corpora (LOB, BNC, BE06) on which the New-GSL were based, represented British-English, and three out of the four corpora (LOB, BE06, EnTenTen12) were made up of written discourse. In the only corpus which included spoken English (BNC), spoken samples accounted for only 10%. Given that the BNC/COCA2000 was developed from a corpus which represents a range of spoken and written discourses and varieties of English, it is understandable why the BNC/COCA2000 is likely to be perceived as more useful by teachers from various contexts and to be learned earlier by L2 learners than the New-GSL. By examining high-frequency word lists from the perspectives of corpus linguistics, teachers, and learners, the present study provides a useful methodological innovation that could be implemented in future research.

One interesting finding of this study is related to the teachers' ratings. The common assumption is that words perceived as being useful for L2 learners may vary greatly between teaching contexts. However, the teacher ratings were less diverse than expected. The teachers in this study came from different L1 backgrounds, had experienced teaching in different EFL/ESL contexts, and varied in years of teaching experience. Yet their ratings were relatively consistent regardless of the criteria of word usefulness and groups of teachers being examined. This finding suggests that L2 teacher perceptions of high-frequency words for L2 learners may be similar across a wide range of contexts. No earlier studies have explored this issue from the perspective of teacher cognition.

This study has an innovative, cross-disciplinary approach towards evaluating high-frequency word lists. It brings together corpus linguistics, Second Language Acquisition, and teacher cognition research under the umbrella of word list studies. While all earlier studies (Brezina & Gablasova, 2015; Gilner & Morales, 2008; Nation, 2004; Nation & Hwang, 1995) used lexical coverage from corpora as the sole criterion, the present study used the information from teachers and learners to supplement corpus-based information in word list evaluation. This approach takes advantage of statistical information to identify and prioritize items that are likely to be encountered by learners. Meanwhile, it ensures that the final entries are appropriate and relevant for L2 learning and teaching. Information from teachers and learners takes into account the contextual and circumstantial realities of a language classroom and provides indicators of the extent to which corpus-based word lists would filter their way to L2 classrooms. As shown in this study, if only corpus information were used in the comparison, it would be challenging to determine whether the BNC/COCA2000 or the New-GSL is a more appropriate for EFL learners in Vietnam. However, when teacher perceptions of word usefulness and learner vocabulary knowledge were used to support corpus-based information, the results provided evidence that the BNC/COCA2000 is the most appropriate list. Additionally, the present study involved the participation of a large number of EFL/ESL teachers and L2

learners. The teachers varied in terms L1 backgrounds, teaching contexts, and years of teaching experience. The learners represented different proficiency levels, university levels, and majors. Furthermore, the information from each source (teachers and learners) was analysed from different angles. A range of criteria were used as indicators of word usefulness and vocabulary knowledge, and the lists were compared as a whole, as well as using the most frequent items in each list. The use of teachers and learners, the large number and great diversity of the participants, and the in-depth analysis of the data provide a thorough assessment of the lists. This provides future studies with a useful model of how to evaluate corpus-based word lists for L2 learning and teaching.

VI Pedagogical implications

The present study indicated that Nation's (2012) BNC/COCA2000 appears to be the most suitable high-frequency word list for EFL learners in Vietnam. This list might also be a useful vocabulary resource for learners in many other EFL contexts. Foreign language learning/teaching in Vietnam shares features of typical foreign language teaching/learning situations described in previous studies (e.g. Webb & Nation, 2017; Muñoz, 2008). In such contexts, learners study English in their home country where English is not the first or significant language and have relatively limited contact with English outside classroom. The time allocated for learning English at school is also limited and the exposure to English during the class periods may be limited in source, quantity and quality. Consequently, the majority of learners in various EFL contexts have insufficient knowledge of high-frequency words despite many years of studying English (e.g. Dang, 2019; Akbarian, 2010; Henriksen & Danelund, 2015; Matthews & Cheng, 2015; Webb & Nation, 2017).

Given the importance of high-frequency words and EFL learners' insufficient knowledge of these words, it is essential for teachers and course designers to ensure that learners have mastered these words before moving on to words at lower frequency levels. At the beginning of a learning program, students' knowledge of the high frequency words should be assessed. Webb, Sasao and Ballance (2017) Updated Vocabulary Levels Test (developed from the first five 1,000 word lists of the BNC/COCA lists) can be used to measure learners' knowledge of the BNC/COCA2000. If the test scores indicate that students have not yet mastered the BNC/COCA2000, teachers should ensure that their classroom programs include a clear focus on learning this vocabulary (for suggestions for school wide programs, see Webb et al., 2017).

There are several important issues teachers should consider when helping students learn the BNC/COCA2000 words. First, there are different aspects involved in knowing a word such as form-meaning relationship, word parts, collocations, and associations (Nation, 2013). Teachers should help learners to gain knowledge of the form-meaning relationship first because it is the most important aspect of vocabulary knowledge, which creates a foundation for the acquisition of other aspects. Once learners have mastered the form-meaning relationship of a word, teachers should create opportunities for them to consolidate and expand knowledge of words.

Second, although this study indicated that the BNC/COCA2000 had more words known by learners and indicated as being useful by teachers, 38 New-GSL words were

known by at least 90% of the learners and indicated as being useful by the teachers. Together with the BNC/COCA2000 words, these 38 New-GSL words should be the set as the learning goal for L2 learners. As vocabulary learning is an incremental process (Schmitt, 2010), the 2,038 words from the BNC/COCA2000 and New-GSL should be considered as a long-term vocabulary learning goal rather than a short-term goal. In its original format, the BNC/COCA2000 is divided into two sub-lists, each of which consists of 1,000 headwords. These sub-lists may be too large to fit in a single course. Therefore, to better assist teachers and learners in setting learning goals, we also rank the BNC/COCA2000 word family headwords according to their mean coverage in nine spoken and nine written corpora (see Appendices 11 and 12 in the Supplementary data). The information about the mean coverage of the 38 New-GSL word family words is already presented in Appendix 10. Teachers and learners can use the information in these appendices to set the short-term vocabulary learning for their courses. One possible way is focusing on words with higher mean coverage before moving on to those with lower mean coverage. Sequencing vocabulary learning in this way would better scaffold learners' vocabulary development, because knowledge of known words would facilitate the acquisition of new words while learning new words would consolidate and expand knowledge of known items (Dang & Webb, 2016; Webb & Nation, 2017).

Third, to create opportunities for learners to learn new words and expanding knowledge of known words, teachers should follow Nation's (2007) Four Strands of meaning-focused input, meaning-focused output, language-focused learning, and fluency development. The four strands provides a framework that includes opportunities to encounter and use words in different contexts. Also, to ensure that students encounter high-frequency words more often, teachers can use vocabulary analysis programs like Lextutor (Cobb, n.d.) or AntwordProfiler (Anthony, n.d.) to check the proportion of BNC/COCA2000 words in the texts and adapt the vocabulary in the texts accordingly.

This study has several limitations which provide avenues for future research. First, only the vocabulary knowledge of Vietnamese EFL learners was examined in the present study. Although these learners share features of learners in many EFL contexts and the data of teachers from different L1 backgrounds with teaching experience in different EFL/ ESL (English as a second language) contexts were used to triangulate the information from the learners, bias towards Vietnamese EFL learners is inevitable. Therefore, while the present study provides useful information about the vocabulary knowledge of a specific learner population, further research with L2 learners in other contexts may provide further insight into knowledge of the BNC/COCA2000 and New-GSL possessed by learners from different L1 populations. Second, this study only measured learners' knowledge of words. It would be useful for future research investigating which words students have learned on their own and which words they have learned from classroom instruction. Third, when measuring the learner participants' vocabulary knowledge, this study does not consider French/English loanwords in Vietnamese. Fourth, it is unlikely that the teacher participants were aware of the two lists during the data collection (between December 2014 and May 2015) because these lists were fairly recent (both of them were available online in 2013) and the name of the word lists were not mentioned to the participants during the recruitment to avoid biasing them toward a certain word list. However, a rigorous follow-up study with all teacher participants after the study completed would

provide more solid evidence about teachers' awareness of the BNC/COCA2000 and the New-GSL list. Fifth, the main purpose of the current study is to compare the BNC/COCA2000 and the New-GSL; therefore, the BNC/COCA2000 words that also appear in the New-GSL were not used as the target words in the teacher surveys and learner vocabulary tests. It would be useful for future research to explore learner knowledge and teacher perception of the usefulness of these items. Such research would provide further insight into the value of the BNC/COCA2000 words for L2 learners. Finally, this study used the word family version of the BNC/COCA2000 for the comparison, and to deal with the limitation that the sub-lists in the original version of the BNC/COCA2000 is too big (1,000 words) to fit in a single course, this study provided list users with the ranking of the BNC/COCA words in terms of the mean coverage of word family headwords. However, given the current trend in word list studies; that is, offering word lists in different formats using different units of counting (Dang, Coxhead, & Webb, 2017; Gardner & Davies, 2014; Nation, 2016), future research on validating high-frequency word lists should be carried out to validate different versions of the lists using different units of counting and rank the items in the lists according to these units of counting.

VII Conclusions

Expanding on Dang and Webb's (2016) study, the present research found that the BNC/COCA2000 had more words known by L2 learners and perceived as being useful by EFL/ESL teachers than the New-GSL. These results suggest that the BNC/COCA2000 is the more suitable high-frequency word list for L2 learners, at least in the Vietnamese EFL context. This study is the first attempt to use information from both teachers and learners to supplement corpus-based information in the evaluation of lists of high-frequency words.

Funding

ORCID iDs

Thi Ngoc Yen Dang https://orcid.org/0000-0002-3189-7776
Averil Coxhead https://orcid.org/0000-0003-3392-6961

Supplemental material

Supplemental material for this article is available online.

Notes

1. Originally, 168 students were recruited for the present study. However, 33 students were excluded from the study because they did not complete all the tests. These students were spread across all 21 classes.
2. The original plan was to select the top 100 words, but there were seven words sharing the same ranking.

References

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners'. *System*, *38*, 391–401.

Anderson, R.C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading Language Research*, *2*, 231–256.

Anthony, L. (n.d.). *AntwordProfiler*. Available at: http://www.laurenceanthony.net/antwordprofiler_index.html (accessed February 2020).

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium.

Banister, C. (2016). The academic word list: Exploring teacher practices, attitudes and beliefs through a web-based survey and interviews. *The Journal of Teaching English for Specific and Academic Purposes*, *4*, 309–325.

Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production: Vocabulary Acquisition and Use in L2 French and Italian. *Studies in Second Language Acquisition*, *34*, 269–290.

Beckner, C., Blythe, R., Bybee, J., et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary?: Introducing the New General Service List. *Applied Linguistics*, *36*, 1–22.

Brown, D. (2018). Examining the word family through word lists. *Vocabulary Learning and Instruction*, *7*, 51–65.

Browne, C. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher*, *4*, 13–16.

Browne, C. (2014). A New General Service List: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction*, *3*, 1–10.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. London: Longman.

Cobb, T. (n.d.). *Lextutor*. Available at: http://www.lextutor.ca (accessed February 2020).

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238.

Dang, T.N.Y. (2019). High-frequency words in academic spoken English: Corpora and learners. *ELT Journal*. Electronic publication ahead of print version. Published online: 20 December 2019 (doi: 10.1093/elt/ccz057).

Dang, T.N.Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (ed.), *The Routledge handbook of vocabulary studies*. New York: Routledge.

Dang, T.N.Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*, 959–997.

Dang, T.N.Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76.

Dang, T.N.Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL – International Journal of Applied Linguistics*, *167*, 132–158.

Dang, T.N.Y., Webb, S., & Coxhead, A. (under review). The relationship between lexical coverage, learner knowledge, and teacher perceptions of the usefulness of high-frequency words.

Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. New York: Routledge.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research*. New York: Routledge.

Dunlea, J., Spiby, R., Nguyen, T.N.Q., et al. (2018). *APTIS-VSTEP comparability study: Investigating the usage of two EFL tests in the context of higher education in Vietnam*. British Council validation series No. VS/201 8/001. Hanoi: British Council.

Ellis, N.C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N.C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396.

Engels, L.K. (1968). The fallacy of word counts. *International Review of Applied Linguistics in Language Teaching*, *6*, 213–231.

Francis, W.N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*, 305–327.

Gerami, R.G., & Noordin, N.B.T. (2013). Teacher cognition in foreign language vocabulary teaching: A study of Iranian high school EFL teachers. *Theory and Practice in Language Studies*, *3*, 1531–1545.

Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, *23*, 65–83.

Gilner, L., & Morales, F. (2008). Corpus-based frequency profiling: Migration to a word list based on the British National Corpus. *The Buckingham Journal of Language and Linguistics*, *1*, 41–58.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.

He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*, *53*, 348–371.

Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In P. Pietilä, K. Doró & R. Pipalová (Eds.), *Lexical issues in L2 writing* (pp. 1–27). Newcastle upon Tyne: Cambridge Scholars Publishing.

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*, 785–800.

Hlaváčová, J. (2006). New approach to frequency dictionaries: Czech example. Unpublished paper presented at the 5th International Conference on Language Resources and Evaluation, Genoa, 24–26 May. Available at: http://www.lrec-conf.org/proceedings/lrec2006/pdf/11_pdf.pdf (accessed February 2020).

Juilland, A.G., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. London: Mouton.

Lau, C., & Rao, N. (2013). English vocabulary instruction in six early childhood classrooms in Hong Kong. *Early Child Development and Care*, *183*, 1363–1380.

Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In P.J.L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Palgrave Macmillan.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, *19*, 255–271.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, *59*, 567–587.

Laufer, B., & Nation, I.S.P. (2012). Vocabulary. In S.M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 163–176). London: Routledge.

Lesaux, N.K., Kieffer, M.J., Faller, S.E., & Kelley, J.G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, *45*, 196–228.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1–13.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*, 142–154.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mochida, K., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*, 73–98.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*, 578–596.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

Nation, I.S.P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*, 1–12.

Nation, I.S.P. (2012). *The BNC/COCA word family lists*. Available at: http://www.victoria.ac.nz/lals/about/staff/paul-nation (accessed February 2020).

Nation, I.S.P. (2013). *Learning vocabulary in another language*. 2nd edition. Cambridge: Cambridge University Press.

Nation, I.S.P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.

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: Cambridge University Press.

Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.

Nation, P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Amsterdam: John Benjamins.

Nation, P., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, *23*, 35–41.

Nguyen, T.M.H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*, 298–320.

Pinchbeck, G.G. (2014, March). Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English. Unpublished paper presented at the Roundtable presentation at American Association of Applied Linguistics, Portland, OR, USA.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Savický, P., & Hlaváčcová, J. (2002). Measures of word commonness. *Journal of Quantitative Linguistics*, *9*, 215–231.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.

Schmitt, N, Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*, 26–43.

Schmitt, N, Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*, 487–512.

Stein, G. (2017). Some Thoughts on the issue of core vocabularies: A response to Vaclav Brezina and Dana Gablasova: 'Is there a core general vocabulary?' Introducing the New General Service List. *Applied Linguistics*, *38*, 759–763.

Stæhr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*, 139–152.

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, *18*, 299–313.

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing*, *22*, 993–1019.

van Heuven, W.J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.

Webb, S.A., & Chang, A.C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*, 113–126.

Webb, S., & Nation, I.S.P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Webb, S., & Rodgers, M.P.H. (2009). The lexical coverage of movies. *Applied Linguistics*, *30*, 407–427.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *IJAL – International Journal of Applied Linguistics*, *168*, 34–70.

West, M. (1953). *A general service list of English words*. London: Longman, Green.

Zhang, W. (2008). In search of English as a foreign language (EFL) teachers' knowledge of vocabulary instruction. Unpublished PhD thesis, Georgia State University, Atlanta, GA, USA.