White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/174018/

Version: Accepted Version

Article:
Dang, TNY, Coxhead, A and Webb, S (2021) Vocabulary in academic spoken English. New Zealand Studies in Applied Linguistics, 26 (2). ISSN: 1173-5562

This item is protected by copyright, all rights reserved. This is an author produced version of an article published in New Zealand Studies in Applied Linguistics. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

VOCABULARY IN ACADEMIC SPOKEN ENGLISH

Thi Ngoc Yen Dang

University of Leeds, U.K

Averil Coxhead

Victoria University of Wellington, New Zealand

Stuart Webb

University of Western, Ontario, Canada

Abstract

Understanding academic spoken English is an important but challenging task for many users of English as an additional language. Vocabulary knowledge plays a significant role in enhancing comprehension, but little is known about the nature of vocabulary in academic spoken English. This article reviews our current studies which are among the very few attempts to address this research gap. It focuses on (a) the number of words needed to comprehend academic spoken English, (b) English for Academic Purposes (EAP) learners' knowledge of high-frequency words, (c) the extent to which academic written word lists cover the vocabulary in academic spoken English, (d) the most frequent and wide ranging words in academic spoken English, and (e) how to incorporate these words in vocabulary learning programs for EAP learners. Directions for future research are also discussed in the article.

Key words: vocabulary; corpus linguistics; academic spoken English; English for Academic Purposes

Introduction

Comprehending academic speech such as lectures, seminars, lab sessions, and tutorials is an essential but challenging task for users of English as an additional language in English medium university programs (Flowerdew & Miller, 1992; Mulligan & Kirkpatrick, 2000). One of the biggest reasons for this challenge is a lack of vocabulary knowledge (Berman & Cheng, 2001; Flowerdew & Miller, 1992). Therefore, to help these learners to enhance their comprehension of academic spoken English, it is crucial to examine the nature of vocabulary in academic speech. Despite this need, studies investigating vocabulary in academic spoken English are very limited in number. Let us take research on academic word lists as an example. A number of academic written word lists have been developed such as Campion and Elley's (1971) Academic vocabulary list, Praninskas's (1972) American university word list, Lynn's (1973) academic word list, Ghadessy's (1979) academic word list, Xue and Nation's (1984) University Word List, Coxhead's (2000) Academic Word List, Gardner and Davies's (2014) Academic Vocabulary List, and Browne, Culligan, & Phillip's (n.d.) New Academic Word List. In contrast, hardly any academic spoken word lists have been created. This article discusses this gap in research.

According to Dang, Coxhead, and Webb (2017), one possible reason why vocabulary in academic spoken English is an underexplored research area is the challenge of creating a large and representative corpus of academic spoken English. Collecting and transcribing spoken data are much more difficult and time consuming than collecting written data (O'Keeffe, McCarthy, & Carter, 2007). For example, while we have academic written corpora of hundreds of millions of words such as the Academic section in the Corpus of Contemporary American English (COCA), the two largest available academic spoken corpora, the British Academic Spoken English corpus (BASE) and the Michigan Corpus of Academic Spoken English (MICASE), are relatively small at around 1.6 million words.

Taken together, it is important to understand the nature of vocabulary in academic spoken English, but not many studies have investigated this area. In this article, we will review our research on vocabulary in academic spoken English, which is among the very few attempts to address this research gap. The article will be organized around five key questions related to vocabulary in academic spoken English:

1. How many words are needed to understand academic spoken English?

2. To what extent do English for Academic Purposes (EAP) learners know high-frequency words?
3. To what extent do existing academic written word lists cover the vocabulary in academic spoken English?
4. What are the most frequent and wide ranging words in academic spoken English?
5. How could the list of these words be used in EAP programs?

How many words are needed to understand academic spoken English?

In order to help users of English as an additional language to improve their comprehension of academic spoken English, it is important to determine the number of words that they need to know to comprehend academic speech. A typical way to address this question is to estimate the number of words needed to reach 95% coverage of academic spoken English. Research (Durbahn, Rodgers, and Peters, 2020; van Zeeland & Schmitt, 2013) found a strong relationship between the percentage of known words in a text (lexical coverage) and listening comprehension. They suggested that although the higher lexical coverage, the better comprehension is, 95% is the point at which reasonable listening comprehension is likely to be achieved.

Corpus-driven research has revealed that the lexical demands of academic spoken English varies according to the specific kinds of academic speech. Dang and Webb (2014) analyzed the vocabulary in the BASE corpus and found that 4,000 word families plus proper nouns (e.g., *Peter, Mary*) and marginal words (e.g., *umh, oh*) are needed to reach 95% coverage of academic lectures and seminars. A word family is made up of a base word together with its inflected forms and derivational forms up to Level 6 in Bauer and Nation's (1993) taxonomy of affixation. For example, the word family *remember* consists of *remember, remembers, remembered, remembering, remembrance, and remembrances*. Subsequent corpus-driven studies, however, have revealed that 3,000 word families plus marginal words and proper nouns are sufficient to achieve 95% coverage of lab sessions and tutorials (Coxhead, Dang & Mukai, 2017), and conference presentations (Dang, under review). These findings suggest that lectures and seminars are probably more demanding than labs sessions, tutorials, and conference presentations in terms of vocabulary size necessary for comprehension. As users of English as an

additional language are expected to engage in various kinds of academic speech, it is likely that they would need to know at least 4,000 word families. In a study on academic spoken English at secondary school, Coxhead (2017) found that teacher talk in classrooms for Grade Six (10 and 11 year old students) in an international school in Germany in Maths, Science and English as an Additional Language (EAL) had quite similar lexical requirements at 95% plus proper nouns and marginal words at 2,000 for EAL and Maths and 3,000 for Science. At 98%, the picture changed quite dramatically, to Science at 7,000, Maths at 6,000 and EAL at 4,000.

To what extent do EAP learners know high-frequency words?

High-frequency words are the words that occur frequently in everyday language such as *cook, go,* and *beautiful* (Nation, 2013).The most frequent 2,000 words (Nation, 2013) or 3,000 words (Schmitt & Schmitt, 2014) of general vocabulary have been widely accepted as high-frequency vocabulary. Learning a small number of high-frequency words would allow users of English as an additional language to recognize a large number of words in various contexts, which will then help to increase their comprehension significantly (Nation, 2006). For example, Dang and Webb (2020) analyzed the vocabulary in 18 corpora which represented different kinds of spoken and written discourse and found that the most frequent 2,000 words in Nation's (2012) BNC/COCA lists covered from75% to 91.60% of the words in these corpora.

Given the benefit of high-frequency words, these words have been reported to be the crucial starting point of vocabulary learning for users of English as an additional language (Nation, 2013; Webb & Nation, 2017). As a result, it is commonly assumed that these learners should already have learned high-frequency words when starting their EAP study (Coxhead, 2000). However, recent studies with EAP university students in different contexts have revealed that this assumption does not always hold true. Akbarian's (2010) study with EAP students in Iran showed that 76% of the students had failed to master the most frequent 2,000 word families. Similarly, Matthews and Cheng (2015) reported that their EAP students in China knew only 77% of the most frequent 2,000 word families. Dang's (2020a) study with EAP students in Vietnam also revealed that only nearly 20% of the participants had mastered the most frequent 2,000 word families, nearly 60% had mastered the most frequent 1,000 word families but no other word frequency levels, and more than 20% had not even mastered the most frequent 1,000 word

families. Drummond (to appear) measured the vocabulary of international students in a pre-sessional university program in the U.K and found that they only knew 79.45% of the words at the 2nd 1,000 word level and 64.89% of the words at the 3rd 1,000 word level. It should be noted that Akbarian (2010), Matthews and Cheng (2015), Dang (2020a), and Drummond (to appear) all measured the receptive knowledge of form-meaning relationship; Matthews and Cheng (2015) focused on spoken forms and other studies focused on written forms. Form-meaning relationship is one of the most important aspect of vocabulary knowledge because it is central to comprehension (Webb & Chang, 2012). However, knowing a word involves many other aspects such as collocation and word association (Nation, 2013, 2020). Therefore, it is fair to say that the participants' vocabulary knowledge of other aspects of vocabulary in the aforementioned studies may be even lower.

Corpus-driven research (Dang & Webb, 2014) has indicated that 4,000 word families are likely to be needed to achieve reasonable comprehension of academic speech. However, research measuring EAP learners' vocabulary knowledge revealed that a reasonable number of these learners may have insufficient knowledge of the most frequent 2,000 words and 3,000 words, let alone the most frequent 4,000 words. These findings help to explain why comprehending academic spoken English is challenging for many users of English as an additional language (Flowerdew & Miller, 1992; Mulligan & Kirkpatrick, 2000) and there is a need for support from EAP programs to help these learners narrow the gaps in their vocabulary knowledge. In the next section, we will examine the extent to which existing academic word lists may potentially help users of English as an additional language to deal with the lexical demands of academic spoken English.

To what extent do existing academic written word lists cover the vocabulary in academic spoken English?

Corpus-driven research has found that the coverage of academic written word lists in academic spoken English is much lower than their coverage in academic written English. Let us take Coxhead's (2000) Academic Word List, which represents the most frequent and wide ranging words in academic written English, as an example. Table 1 shows that this list consistently covers just over 4% of different kinds of academic speech. These coverage figures are much

lower than its coverage in academic written texts (10%) (Coxhead, 2011, 2016). This suggests that vocabulary in academic spoken English may be different from that in academic written English and that knowledge of items in academic written word lists is probably insufficient to help EAP learners to deal with vocabulary in academic spoken English. This leads to the need for developing an academic spoken word list.

Table 1. Coverage of Coxhead's (2000) Academic Word List in academic spoken English

Kinds of speech events	Coverage
Lectures (Thompson, 2006)	4.6%
Lectures & seminars (Dang & Webb, 2014)	4.41%
Lectures, seminars, labs, and tutorials (Coxhead et al., 2017)	4.17%
Tutorials (Coxhead & Dang, 2019)	3.56%
Labs (Coxhead & Dang, 2019)	2.52%

Another issue with existing academic word lists is that they assume EAP learners to be a homogeneous group that has the same vocabulary knowledge. For example, Dang's (2000) Academic Word List was developed with the assumption that learners already know the most frequent 2,000 word families of general English. However, as shown by Akbarian (2010), Matthews and Cheng (2015), Dang (2020a), and Drummond (forthcoming), the vocabulary knowledge of EAP learners varies. While some learners manage to master the most frequent 2,000 words or more, others struggle to achieve this goal. Word lists should match the proficiency level of their users (Nation, 2016). The variation in EAP learners' vocabulary knowledge leads to the need for an academic spoken word list that is adaptable to learner proficiency levels.

In recognition of these needs, Dang et al. (2017) developed an Academic Spoken Word List which (a) captures the most frequent and wide ranging words of academic spoken English and (b) is adaptable to list users' proficiency levels. In the next section, we will describe the list in more detail.

What are the most frequent and wide ranging word families in academic spoken English?

The Academic Spoken Word List aims to serve the needs of learners in English for General Academic Purposes (EGAP) programs; that is, the programs in which (a) there is a mixture of students planning to study various subject areas, (b) learners are unclear about their target subject areas, (c) teachers lack background knowledge of learners' specific disciplines, and (d) in interdisciplinary environments where it is unclear which specific discipline an academic subject belongs to. The Academic Spoken Word List was developed from an academic spoken corpus of 13 million running words. The corpus represents four kinds of speech events: lectures, seminars, labs, and tutorials and at least 7 varieties of English.

Table 2 presents the component of the corpus from which the Academic Spoken Word List was developed. As can be seen from the table, this corpus is divided into four disciplinary sub-corpora based on Becher's (1989) classification of academic disciplines in higher education: hard pure (e.g., *biology, mathematics, physics*), hard applied (e.g., *computer sciences, health and medical sciences, mechanical engineering*), soft pure (e.g., *philosophy, history, art*), and soft applied (e.g., *business, law, education*). Each sub-corpus contains 3.25 million running words and represents six subject areas. Each subject area has around 500,000 words. Given the size and structure of the corpus, it is expected that the corpus can represent as closely as possible the words that students are likely to encounter in speech during their academic studies.

A total of 1,741 word families were selected from the academic spoken corpus to include in the Academic Spoken Word List. The word family rather than the lemma was chosen as the unit of counting of the Academic Spoken Word List for several reasons. First, following Coxhead (2000) and Nation (2013), Dang et al. (2017) assume that learners pick up knowledge of word family members during their learning process, and they are provided with training on word part knowledge and word building skills. Second, studies investigating learners' derivational knowledge (e.g., Mochizuki & Aizawa, 2000; Sasao & Webb, 2017) has found that even beginner learners already known some closely related derivations (e.g., *-full, un-*). Choosing the lemma as the unit of counting may overestimate the learning burden of the words from the Academic Spoken Word List. Third, Dang et al. aimed to integrate the Academic Spoken Word List with Nation's (2012) BNC/COCA lists in a systematic program to enhance learners' comprehension of academic spoken English. The word family was chosen as the unit of counting

of the BNC/COCA list for consistency because it is the unit of counting of the BNC/COCA lists (Please see Dang et al. (2017) for more details). To be selected, these word families needed to meet the range (i.e. occurring in all four sub-corpora and at least 50% of the subject areas), frequency (i.e. having relative frequency of at least 26.9 times per million words) and dispersion (i.e. having a Julliand's D of at least 0.6.) criteria. Range, frequency, and dispersion are important criteria to select items for corpus-driven word lists (Dang, 2020b; Nation, 2016). These criteria ensured that the ASWL included the most frequent and wide ranging words in academic spoken English.

Table 2. Component of the corpus used to develop the Academic Spoken Word List

Sub-corpora	Subject	Size
Hard-pure (3,261,623 words)	Astronomy	593,062
	Biology	552,452
	Chemistry	556,138
	Ecology & Geology	555,312
	Mathematics	450,481
	Physics	554,178
Hard applied (3,254,094 words)	Chemical Engineering	563,938
	Computer Sciences	555,175
	Cybernetics	555,401
	Electrical Engineering	550,181
	Health & Medical Sciences	470,795
	Mechanical Engineering	558,604
Soft pure (3,256,283 words)	Art	553,160
	Cultural Studies	498,393
	History	554,214
	Philosophy	549,577
	Political Studies	545,059
	Psychology	555,880
Soft applied (3,257,661 words)	Business	513,133
	Economics	610,998
	Education	571,023
	Law	616,398
	Management	461,093
	Public Policy	485,016

The Academic Spoken Word List covered 90.13% of the corpus from which it was developed. It provided roughly the same amount of coverage in each sub-corpus: 89% (soft pure), 89.46% (hard pure), 90.92% (soft applied), and 91.07% (hard applied). This indicates learners from different academic disciplinary groups might gain fairly similar benefit from the Academic Spoken Word List in terms of lexical coverage. The list was further validated in three independent corpora: (a) another academic spoken corpus of similar size and structure (the second academic spoken corpus), (b) an academic written corpus, and (c) a non-academic spoken corpus. To ensure a fair comparison, these corpora had the same size as the corpus used to develop the Academic Spoken Word List (around 13 million words). The validation showed that the coverage of the Academic Spoken Word List in each academic spoken corpus was roughly the same (around 90%), which was higher than its coverage in the non-academic spoken corpus (87.06%) and the academic written corpus (81.81%). This suggests that the list better represents academic than non-academic vocabulary, and better represents spoken than written vocabulary.

One innovative feature of the Academic Spoken Word List is that it is adaptable to learners' current vocabulary levels. The Academic Spoken Word List was divided into four levels based on Nation's (2012) BNC/COCA lists (see Table 3). Words at Levels 1, 2, and 3 of the Academic Spoken Word List are also words at the 1st, 2nd, and 3rd 1,000 BNC/COCA lists while words at Level 4 of the Academic Spoken Word List are outside the most frequent 3,000 words.

Table 3. The four levels of the Academic Spoken Word List (ASWL) (Dang et al., 2017, p. 979)

ASWL level	BNC/COCA level	Number of word families	Examples
Level 1	1st 1,000	830	*alright, know, stuff*
Level 2	2nd 1000	456	*therefore, determine, approach*
Level 3	3rd 1000	380	*achieve, significant, aspect*
Level 4	4th 1000 onwards	75	*arbitrary, optimize, theorem*

Depending on their current knowledge of general vocabulary, learners can skip certain levels of the Academic Spoken Word List. Learners with knowledge of the most frequent 2,000 and 3,000 word families only need to learn 455 word families or 75 words families from the Academic Spoken Word List that are beyond their current vocabulary levels, respectively. That knowledge might enable them to recognize around 95% of the word in academic spoken English. This is

encouraging because to achieve 95% coverage of academic spoken English (lectures & seminars), 4,000 word families of general vocabulary is needed (Dang & Webb, 2014). This means if learners with knowledge of the most frequent 2,000 word families and 3,000 word families do not want to learn items from the Academic Spoken Word List, they might need to learn an extra of 1,000-2,000 word families of general vocabulary to reach 95% coverage of academic spoken English. Thus, learning the Academic Spoken Word List provides a useful shortcut to understanding academic speech.

As for learners who have mastered only the most frequent 1,000 words or fewer, ideally they would study the most frequent 2,000 and even 3,000 BNC/COCA word families and then move to the relevant BNC/COCA word level so that they can achieve 95% coverage of academic spoken English. However, this may be a demanding goal for some learners. As shown by Akbarian (2010), Matthews and Cheng (2015), Dang (2020a), and Drummond (to appear), a reasonable proportion of EAP learners in different contexts are failing to master the most frequent 1,000 and 2,000 word families after a long period of studying English. Therefore, learning 3,000 words or more within a short period of time might be too demanding a goal for them. However, focusing on items from the Academic Spoken Word List may help to solve this dilemma to some extent. These learners may need to learn 1,741 words (those having yet to master the most frequent 1,000 words) and 911 words (those having mastered the most frequent 1,000 words) and that knowledge would allow them to recognize 92%-93% of the words in academic spoken English. Importantly, knowledge of the academic spoken word list might also enable learners to deal with 92%-93% of the words in general spoken English.

The Academic Spoken Word List was originally developed with the aim to help learners comprehend lectures, seminars, labs, and tutorials. However, subsequent studies have shown that this list might be a useful vocabulary resource to enhance EAP learners' comprehension of other kinds of academic speech events. Liu and Chen (2019) examined the occurrence of items from the Academic Spoken Word List in a 4.37 million word corpus of TED Talks. They found that the list covered 89.6% of the corpus. Dang (under review) examined the coverage of the list in a 565,758 word corpus of academic conference presentations. She found that the list covered 87.52% of the corpus. If learners' current vocabulary is taken into account, learning items from

the Academic Spoken Word List would help learners to achieve a potential coverage of 95% to 97% of academic conference presentations.

Taken together, research on the Academic Spoken Word List has provided evidence that this list is a useful resource to help EAP students from various disciplines and vocabulary levels enhance their comprehension of academic lectures, seminars, lab sessions, tutorials, academic conference presentations, and TED talks. The Academic Spoken Word List can be freely downloaded from https://osf.io/gwk45/

How could the Academic Spoken Word List be used in EAP programs?

It should be noted that creating the Academic Spoken Word List does not mean that we should dismiss existing academic word lists such as Coxhead's (2000) Academic Word List and Gardner and Davies's (2014) Academic Vocabulary List. In fact, the Academic Spoken Word List should be used together with academic written word lists to support students' reading and listening comprehension in EAP programs. One strength of the Academic Spoken Word List is its adaptability to learners' vocabulary levels. This feature offers teachers and program designers flexibility to incorporate the ASWL into their EAP programs to match their students' needs. Let us explain this idea in more detail. At the beginning of the program, teachers should measure their students' levels of general vocabulary using vocabulary levels tests such as the Listening Vocabulary Levels Test (McLean, Kramer & Beglar, 2015), the Partial Dictation Test (Matthews & Cheng, 2015), or the Updated Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017). Based on their students' level of general vocabulary and the specific teaching context, teachers can use Figure 1 as a guide to determine the suitable learning sequence and learning goals for their students. For example, if students have yet to master the most frequent 1,000 words and would like to focus on the words that are frequent in academic spoken English, they can start learning items from Level 1 of the Academic Spoken Word List and move gradually to items in Levels 2, 3, and 4. However, if these students would like to broaden their knowledge of general vocabulary first, they could start learning items from Nation's (2012) BNC/COCA lists first. Once they think that their knowledge of general vocabulary is sufficient, they can move on to the Academic Spoken Word List level that is beyond their current vocabulary knowledge. For instance, once learners have learned the 1st and 2nd 1,000 BNC/COCA words, they may skip

Levels 1 and 2 of the Academic Spoken Word List and can focus on Level 3 instead. Once teachers and course designers have identified the learning goals and sequence for their students, they should follow Nation's (2007) Four Strands principles to design activities for students to encounter and use the target lexical items in a meaningful way.

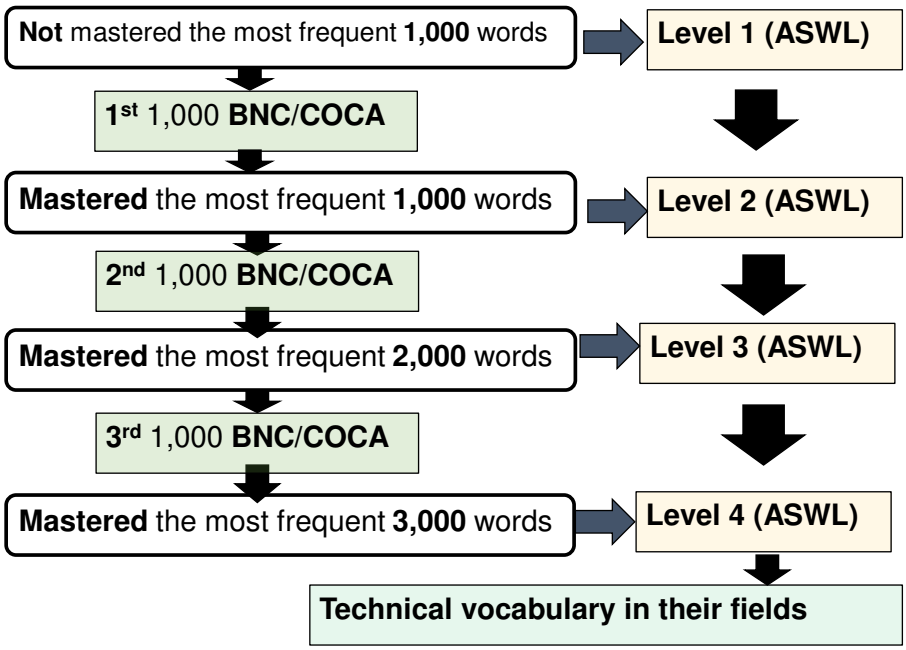

Figure1. Vocabulary learning sequences (Dang et al., 2017, p. 987)

Limitations and future research

The Academic Spoken Word List should be an essential source in English for General Academic Purposes courses. However, in English for Specific Academic Purposes (ESAP) programs or English for Specific Purposes (ESP) programs, discipline-specific word lists may be more relevant than general academic word lists (Dang, 2018a; Hyland & Tse, 2007). Several word lists have been created such as Dang's (2018a) Hard Science Spoken Word List, Dang's (2018b) Soft Science Spoken Word List, and Dang's (2020c) Medical Spoken Word List. However, more lists may need to be developed to reveal the specialized vocabulary in each subject area and to meet the needs of students in a specific academic discipline. Moreover, in academic listening, learners need to process a large amount of dense and abstract information under the time pressure (Biber, 2006), and they have to deal with the challenges caused by the features of spoken English (e.g.,

connected speech, speakers' speech rate and accents, mismatch between spoken and written forms) (Field, 2011; Flowerdew, 1994; Goh, 2000). Therefore, knowledge of multiword units is also important for comprehension of academic spoken English (Simpson-Vlach & Ellis, 2010; Siyanova-Chanturia & Pellicer-Sánchez, 2019). Several lists of lexical bundles in academic speech have been developed (Biber, Conrad, & Cortes, 2004; Coxhead et al., 2017; Simpson-Vlach & Ellis, 2010), but further research on multiword units in academic spoken English is warranted. Additionally, the Academic Spoken Word List was created based on the information from corpora. To make the list better meet the needs of EAP learners and teachers, further validation with teachers and learners is necessary. It would be useful to explore the effectiveness of the learning sequence in Figure 1 in an actual EAP classroom. Furthermore, most previous research on EAP learners' vocabulary knowledge used a test of knowledge of written forms and meanings. Scores on aural vocabulary tests better correlated with listening comprehension than scores on written vocabulary tests (Milton, Wade & Hopkins, 2010). Although several tests have been developed to measure knowledge of spoken vocabulary, they are either not publically available (Aural_Lex Yes/No test) or validated with learners from a specific L1 (Listening Vocabulary Levels Test, Partial Dictation Test). This is certainly an area that needs further investigation. Last but not least, similar to research on academic vocabulary, studies on academic spoken vocabulary has mainly focused on university contexts. More research on academic spoken English in secondary school contexts is needed.

Conclusion

This article reviews our current research on vocabulary in academic spoken English. It shows that knowledge of 4,000 word families is needed to achieve reasonable comprehension of academic spoken English. Meeting this vocabulary goal, however, is challenging for many EAP learners. An academic spoken word list of 1,741 word families was developed to help learners deal with this challenge. The list provides EAP learners with a short cut to enhance their comprehension of academic spoken English. It benefits EAP learners irrespective of their disciplines and vocabulary levels and is a useful resource for setting learning goals and sequences and designing courses and materials for EAP programs. The Academic Spoken Word List provides a foundation for further investigation into academic spoken vocabulary.

References

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, *38*(3), 391–401. https://doi.org/10.1016/j.system.2010.06.013

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Becher, T. (1989). *Academic tribes and territories*. The Society for Research into Higher Education and Open University Press.

Berman, R., & Cheng, L. (2001). English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics*, *4*(1–2), 25–40.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in University teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Browne, C., Culligan, B., & Phillips, J. (n.d.). A new academic word list. Retrieved from http://www.newacademicwordlist.org/

Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. New Zealand Council for Educational Research.

Coxhead, A. (2017). Academic vocabulary in teacher talk: Challenges and opportunities for pedagogy. *Oslo Studies in Language*, *9*(3), 29–44. https://doi.org/10.5617/osla.5845

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, *45*(2), 355–362. https://doi.org/10.5054/tq.2011.254528

Coxhead, A. (2016). Reflecting on Coxhead (2000), “A New Academic Word List". *TESOL Quarterly*, *50*(1), 181–185. 10.1002/tesq.287

Coxhead, A, & Dang, T. N. Y. (2019). Vocabulary in university tutorials and laboratories. In K. Hyland & L. Wong (Eds.), *Specialised English: New directions in ESP and EAP research and practice* (pp. 120–134). Routledge.

Coxhead, A, Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, *30*, 66–78. https://doi.org/10.1016/j.jeap.2017.11.001

Dang, T. N. Y. (2018a). A Hard Science Spoken Word List. *ITL - International Journal of Applied Linguistics*, *169*(1), 44–71. https://doi.org/10.1075/itl.00006.dan

Dang, T. N. Y. (2018b). The nature of vocabulary in academ ic speech of hard and soft sciences. *English for Specific Purposes*, *51*, 69–83. https://doi.org/10.1016/j.esp.2018.03.004

Dang, T. N. Y. (2020a). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 288–304). Routledge.

Dang, T. N. Y. (2020b). High-frequency words in academic spoken English: Corpora and learners. *ELT Journal*, *74*(2), 146–155. https://doi.org/10.1093/elt/ccz057

Dang, T. N. Y. (2020c). The potential for learning specialized vocabulary of university lectures and seminars through watching disciplines-related TV programs: Insights from medical corpora. *TESOL Quarterly*, *54*(2), 436–459. https://doi.org/10.1002/tesq.552

Dang, T. N. Y. (under review). The lexical profile of academic conference presentations. *English for Specific Purposes.*

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. 10.1111/lang.12253

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Dang, T. N. Y., & Webb, S. (2020). Vocabulary instruction and the good language teachers. In C. Griffiths, Z. Tajeddin, & A. Brown (Eds.), *Lessons from good language teachers* (pp. 203–218). Cambridge University Press. https://doi.org/10.1017/9781108774390

Drummond, A. (to appear). Chinese speakers' vocabulary size: Correlations with IELTS scores andimplications for textual coverage.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 1–13. https://doi.org/10.1016/j.system.2019.102166

Field, J. (2011). Into the mind of the academic listener. *Jounal of English for Academic Purposes*, *10*, 102–112. https://doi.org/10.1016/j.jeap.2011.04.002

Flowerdew, J. (1994). *Academic listening: Research perspectives*. Cambridge University Press.

Flowerdew, J., & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal*, *23*(2), 60–80. https://doi.org/10.1177/003368829202300205

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1177/003368829202300205

Ghadessy, P. (1979). Frequency counts, words lists, and materials preparation: A new approach. *English Teaching Forum*, *17*(1), 24–27.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Hyland, K, & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Liu, C. Y., & Chen, H.-H.-J. (2019). Academic spoken vocabulary in TED Talks: Implications for academic listening. *English Teaching & Learning*, *43*, 353–368.

Lynn, R. W. (1973). Preparing word-lists: A suggested method. *RELC Journal*, *4*(1), 25–28. https://doi.org/10.1177/003368827300400103

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*(6), 741–760. https://doi.org/10.1177/1362168814567889

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence inEnglish as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83–98). Multilingual Matters.

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, *28*(2), 291–304. https://doi.org/10.1016/S0346-251X(00)00013-0

Mulligan, D., & Kirkpatrick, A. (2000). How much do they understand? Lectures, students and comprehension. *Higher Education Research & Development*, *19*(3), 311–335. https://doi.org/10.1080/758484352

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82.

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

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

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins. https://doi.org/10.1075/z.208

Nation, I. S. P. (2020). The different aspects of vocabulary knowledge. In S Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15–29). Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511497650

Praninskas, J. (1972). *American university word list*. Longman.

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, *21*(1), 12–30. https://doi.org/10.1177/1362168815586083

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Simpson-Vlach, R., & Ellis, N. (2010). An Academic Formulas List: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Siyannova-Chanturia, A., & Pellicer-Sánchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. Routledge.

Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In Ken Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253–270). Peter Lang.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113–126. 10.1177/0033688212439367

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

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, *168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, *3*(2), 215–229.