



UNIVERSITY OF LEEDS

This is a repository copy of *A corpus-based study of vocabulary in conference presentations*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/193582/>

Version: Accepted Version

Article:

Dang, TNY orcid.org/0000-0002-3189-7776 (2022) A corpus-based study of vocabulary in conference presentations. *Journal of English for Academic Purposes*, 59. 101144. ISSN 1475-1585

<https://doi.org/10.1016/j.jeap.2022.101144>

© 2022 Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

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.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A corpus-based study of vocabulary in conference presentations

Abstract

Understanding conference presentations is an important but challenging task for EAP learners in their academic careers. Given the importance of vocabulary knowledge for comprehension, this study examines (a) the lexical demands of conference presentations, (b) the coverage of the Academic Spoken Word List (ASWL) (Author *et al.*, XXXXa) in these presentations, and (c) the extent to which this list can help EAP learners deal with the lexical demands of these conferences. A 565,758-word corpus was created from conference presentations of 20 academic subject areas. Analysis revealed that together with proper nouns and marginal words, the most frequent 3,000 word families of general English covered 96.84% and the most frequent 5,000 word families covered 98.22% of conference presentations. The ASWL itself covered 87.52% of the academic conference presentation corpus and more than 87% of its sub-corpora. If learners' prior knowledge of general vocabulary is taken into account, learning the ASWL word families may help users to achieve potential coverage from nearly 95% to 97% of conference presentations. Implications on how to transfer the findings to teaching will also be discussed.

Key words: vocabulary; conference presentation; lexical coverage; listening comprehension; Academic Spoken Word List

1. Introduction

Research excellence is a key criterion in evaluating university performance internationally. Therefore, many higher education institutions encourage their research students and scholars to attend and exchange knowledge at international conferences (Forey & Feng, 2016; Morell, 2015). These conferences are great opportunities for them to be updated with the current

development in their fields and broaden their professional network (Feak, 2013). However, as English is usually chosen as the means of communication in most conferences, understanding conference presentations may be challenging for a number of attendees whose L1 is not English (Morell, 2015; Pérez-Llantada et al., 2011), especially those who “have learned English in an idealized, perhaps written, form and have thus not been exposed to the characteristics of rapid colloquial speech” (Flowerdew, 1994, p. 10). This may then make it difficult to get the most out of a conference.

Although various factors contribute to listening comprehension (Goh, 2013), vocabulary knowledge is an important factor (van Zeeland & Schmitt, 2013; Wallace, 2020). To support research students and scholars whose L1 is not English in comprehending conference presentations, it is therefore important to determine (a) the lexical demands of conference presentations and (b) the extent to which existing word lists can help users to deal with these demands. Despite this need, vocabulary in conference presentations is an underexplored area of both ESP research and vocabulary research. ESP research on conference presentations has mainly focused on the rhetorical structure (Rowley-Jolivet & Carter-Thomas, 2005; Shalom, 1993), the use of spoken, written, and visual modes (Forey & Feng, 2016; Morell, 2015), discourse markers (Fernández-Polo, 2014), and humor (Reershemius, 2012). Meanwhile, research on academic vocabulary has mainly examined the lexical demands of written texts. The few studies investigating spoken discourse have focused on vocabulary in university lectures (Author *et al.*, XXXXb; Thompson, 2006), seminars (Author *et al.*, XXXXb), labs (Author *et al.*, XXXXc), and tutorials (Author *et al.*, XXXXc). Little has been done to explore the lexical demands of conference presentations. Similarly, most word lists have been developed to enhance comprehension of academic written English (e.g., Coxhead, 2000; Gardner & Davies, 2014; Paquot, 2010). Only the Academic Spoken Word List (Author et al, XXXXa) has focused on vocabulary in academic spoken discourse. Yet this list was developed to support comprehension of lectures, seminars, labs, and tutorials. The extent to which it can support comprehension of non-class academic speech events such as conference presentations is unclear.

The current study examines (a) the lexical demands of conference presentations and (b) the percentage of words in conference presentations that list users with different vocabulary levels may recognize with the support of the Academic Spoken Word List. The study provides a clearer

idea about the lexical challenge faced by research students and scholars whose L1 is not English when trying to understand conference presentations in English and the value of the Academic Spoken Word List as an instrument to help them deal with this challenge.

1.1. The number of words needed for comprehension of academic spoken English

Vocabulary researchers (e.g., Schmitt et al., 2011) have proved that the larger percentage of known words in a text, the better comprehension is. The percentage of known words in a text is called lexical coverage (Nation & Waring, 1997). While other factors may influence comprehension, lexical coverage is probably the most important factor (Laufer & Sim, 1985). For this reason, a common approach towards identifying the number of words needed for comprehension of text is to estimate the vocabulary size necessary to achieve certain lexical coverage points (Nation & Webb, 2011). There is no lexical coverage threshold for listening comprehension, but higher coverage is likely to result in better comprehension (Schmitt et al., 2011). van Zeeland and Schmitt (2013, p. 474) found that 90%, 95%, and 98% coverage is likely to lead to ‘relatively high’, ‘relatively high and stable’ and ‘very high’ listening comprehension rates, respectively. As 95% and 98% result in higher levels of comprehension, they are usually adopted as the lexical coverage cut-off points to indicate comprehension of texts (van Zeeland & Schmitt, 2013; Webb & Rodgers, 2009a, 2009b).

Previous research on the lexical demands of academic speech has found that a vocabulary size of 4,000 word families and 8,000 word families was needed to achieve 95% and 98% coverage of lectures and seminars, respectively (Author *et al.*, XXXXb). In contrast, knowledge of 3,000 word families may allow learners to achieve 95% coverage of labs and tutorials while knowledge of 7,000 word families and 4,000 word families may enable them to reach 98% coverage of labs and tutorials, respectively (Author *et al.*, XXXXc). A word family (e.g., *present*) consists of a base form (e.g., *present*), its inflected forms (e.g., *presents*, *presented*, *presenting*), and closely related derived forms (e.g., *presentable*, *presentably*, *presentation*, *presentations*, *presentational*, *presenter*, *presenters*, *unpresentable*). As the lexical demands of academic speech vary according to the type of speech events, findings of research on other academic speech events may not necessarily reflect the nature of vocabulary in conference presentations. Given the importance of conference presentations, research on the lexical demands of this kind of speech events is warranted.

1.2. Coverage of academic word lists in academic spoken English

According to Nation (2013), academic vocabulary is commonly defined as the words that occur frequently in a wide range of academic disciplines. It is different from technical vocabulary, which refers to lexical items that occur frequently in a specific discipline or topic. As academic words cover a reasonable proportion of academic texts, knowledge of these words would help to improve comprehension significantly (Coxhead, 2000). However, research has shown that mastering academic vocabulary is challenging for many L2 learners, even advanced level learners (Pérez-Llantada et al., 2011; Vidal, 2003, 2011). Most of academic words are lexical items supporting the presentation of discipline-specific context rather than being closely related to specific disciplines (Pérez-Llantada et al., 2011; Vidal, 2003, 2011). Therefore, unlike technical words, academic words are not likely to be learned as part of L2 learners' specialization. Moreover, a number of academic words occur frequently in general conversation, but their technical meanings are different from their meanings in general conversation (Author, XXXXa; Coxhead, 2020; Gardner & Davies, 2014; Nation, 2016), which may make these words less salient to learners.

Given the importance and challenge of learning academic vocabulary, a number of word lists have been developed to help L2 learners to improve their comprehension of academic written English such as Coxhead's (2000) Academic Word List and Gardner and Davies' (2014) Academic Vocabulary List. However, research has indicated that learning items from academic written word lists may not be sufficient to understand academic spoken English. For example, Coxhead's (2000) Academic Word List consistently covered around 10% of the words in academic written English (see Coxhead, 2011, 2016 for detailed reviews). Yet this list only accounted for 4.41% of the words in lectures and seminars (Author *et al.*, XXXXb), 3.56% of the words in tutorials (Author *et al.*, XXXXd), 2.52% of the words in lab sessions (Author *et al.*, XXXXd), and 2.4% of the words in student presentations (Hincks, 2003).

In recognition of this gap, Author *et al.* (XXXXa) developed an Academic Spoken Word List (ASWL) of 1,741 word families to help learners deal with vocabulary in university lectures, seminars, labs, and tutorials. The ASWL words appear frequently in a 13-million word corpus of academic lectures, seminars, labs, and tutorials from 24 subject areas (e.g., mathematics, chemistry, computer science, medicine, arts, political studies, business, and law). The ASWL

provided 90.13% coverage of the academic spoken corpus from which it was developed and similar coverage of an independent academic spoken corpus against which it was validated¹. This is meaningful because coverage of 90% is likely to help L2 learners to achieve ‘relatively high’ comprehension of spoken texts (van Zealand & Schmitt, 2013, p. 474). Detailed information about these corpora is available at <https://osf.io/tp687/> .

Although the ASWL was fairly long (1,741 word families), it does not mean that learners need to learn all ASWL words. One innovative feature of the ASWL is its adaptability to the users’ current knowledge of general vocabulary. Previous academic word lists considered learners to have a similar knowledge of general vocabulary when starting learning academic vocabulary. For example, Coxhead’s (2000) Academic Word List was developed with the assumption that learners already knew the most frequent 2,000 word families of general vocabulary and therefore excluded these word families from this list. However, research has shown that learners’ vocabulary knowledge varies when they start learning academic vocabulary (Akbarian, 2010; Author, XXX). Some already know the most frequent 2,000 or even 3,000 word families while others have insufficient knowledge of the most frequent 1,000 or 2,000 word families. To benefit learners with different knowledge of general vocabulary, the ASWL has four levels (see Table 1).

Table 1

Four levels of Author et al.’s (XXXXa) Academic Spoken Word List

ASWL level	BNC/COCA word level	Number of items	Examples
Level 1	1 st 1,000	830	<i>think, sorry, indeed, afraid</i>
Level 2	2 nd 1,000	456	<i>discuss, identify, slide, somewhat</i>
Level 3	3 rd 1,000	380	<i>method, impact, fundamental, chart</i>
Level 4	Outside the most frequent 3,000 words	75	<i>optimise, align, horizontal, dynamic</i>
Total		1,741	

Levels 1, 2, and 3 of the ASWL are word families that are common in academic spoken English but are also among the most frequent 3,000 word families of general English represented by Nation's (2012) 1st, 2nd, and 3rd 1,000 BNC/COCA word levels (see section 2.2 below for further explanation of the BNC/COCA lists). Level 4 of the ASWL are word families that are common in academic spoken English but are not among the most frequent 3,000 word families of general English. The technicality of the list increases from Level 1 to Level 4. Depending on their general vocabulary knowledge, learners may only need to focus on certain levels of the ASWL. For example, if they already know the most frequent 2,000 words of general English, they are probably able to recognize the ASWL words at Level 1 and 2. Therefore, they can skip these levels and focus on 455 items from Levels 3 and 4 of the ASWL. This number of ASWL word families is smaller than the number of Coxhead's (2000) Academic Word List (570 word families).

Sequencing vocabulary learning in this way would ensure that the learning and teaching time is spent effectively. Depending on their current vocabulary knowledge, learning the ASWL may allow list users to recognize 92%-96% of the words in lectures, seminars, labs, and tutorials. For this reason, Author *et al.* (XXXXA) suggest that the ASWL is a useful instrument for L2 learners in EAP programs to deal with the vocabulary in university lectures, seminars, labs, and tutorials. The ASWL and its levels can be downloaded for free from the Open Science Framework at <https://osf.io/gwk45/> or the IRIS digital repository at <http://www.iris-database.org>.

As the ASWL is fairly recent, only Author *et al.* (XXXXd) have examined the coverage of the list in specific kinds of speech events. They found that the list covered 92.35% of the words in tutorials and 90.58% of the words in labs. These coverage figures are fairly consistent with the figure reported by Author *et al.* (XXXXa) (about 90%), which supports their claim that the ASWL is a valuable tool to enhance comprehension of academic speech including labs and tutorials. This list could probably support comprehension of conference presentation as well, and this study would test this hypothesis.

1.3. Research questions

The literature review has revealed that no studies have investigated the lexical demands of conference presentations, nor have any studies explored the extent to which the Academic Spoken Word List (ASWL) can help users to meet these lexical demands and improve their

comprehension of conference presentations. To address these gaps, the current study aims to answer three research questions:

1. What vocabulary size is needed to achieve 95% and 98% coverage of conference presentations?
2. What is the coverage of the ASWL in conference presentations?
3. With knowledge of the ASWL, how much coverage of conference presentations may be achieved by list users with different levels of general vocabulary?

2. Methodology

2.1. Corpus

The conference presentation corpus in this study was created from the transcripts of 104 conference and colloquia presentations organized in various countries. Each presentation includes speakers' presentation and a Q& A session between the speakers and the audiences. This means most of the language in the corpus is from the presenter, but a small proportion is from the audiences in the Q & A sessions. These transcripts were collected from six corpora: the Michigan Corpus of Academic Spoken English (MICASE), the Limerick-Belfast Corpus of Academic Spoken English (LIBEL), the John Swales Conference Corpus (JSCC), the English as a Lingua Franca in Academic Setting (ELFA) corpus, the Hong Kong Corpus of Spoken English (HKCSE), and the Vienna-Oxford International Corpus of English (VOICE). Therefore, it is expected that the conference presentation corpus reflects the diverse contexts of conferences that research students and scholars are likely to be exposed to.

As a whole the corpus consists of 565,758 words. To the best of my knowledge, this is the largest corpus of conference presentations that has ever been created. The corpus has two sub-corpora: hard science subjects (156,006 words) and soft science subjects (409,752 words) (see Table 2). The hard/soft division was chosen to structure the sub-corpora in the present study because previous research (e.g., Biglan, 1973a, 1973b) examined academic disciplines from different angles and suggested that the hard/soft division is the strongest dimension that differentiates academic disciplines in higher education. Hard sciences (e.g., chemistry, medicine) tend to have a single paradigm - "a body of theory which is subscribed by all members of the field" while soft sciences (e.g., economics, philosophy) tend to lack this paradigm (Biglan,

1973a, p. 201). In fact, this division has been used to structure the British Academic Spoken English corpus as well as Author *et al.*'s (XXXXa) Academic Spoken Corpus. Due to the challenge of collecting spoken data, the corpus includes more transcripts from soft science subjects than hard science subjects, which affects the balance of the corpus. Therefore, findings of this study are exploratory.

2.2. Analysis

Heatley *et al.*'s (2002) RANGE program was used for the analysis. This program can identify which words in a list appear in the text. To determine the vocabulary size needed to achieve 95% and 98% coverage of academic conference presentations, Nation's (2012) 25 lists of word families from the 1st to the 25th 1,000 words in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) were used with RANGE to represent the 1,000-word level at which the words in the text occurred. The 25 BNC/COCA lists were developed from the BNC and COCA, the two largest corpora to date which represent a wide range of spoken and written genres (e.g., fiction, newspapers, academic texts, news broadcasts, movies, telephone conversations). Items at the 1st 1,000-word level are the most frequent. Items at the 2nd 1,000-word level are more frequent than those at the 3rd 1,000 word levels, but less frequent than those at the 1st 1,000 word level. The further a certain frequency level moves from the 1,000 word level, the less frequent the word families at that level are.

Apart from the 25 BNC/COCA lists, Nation's (2012) list of proper nouns and list of marginal words were also used in the analysis. Earlier studies examining the lexical demands of spoken texts (e.g., Webb & Rodgers, 2009a, 2009b) considered that proper nouns (e.g., *Martin, London*) and marginal words (e.g., *hmm, ohh*) could support comprehension, but the learning burden of these words is light because learners may recognize them from context. Therefore, they counted the coverage provided by these words as part of the overall coverage. This approach was taken in the present study.

Table 2

Components of the conference corpus (565,758 words)

Hard science subjects	Number of words	Number of presentations	Soft science subjects	Number of words	Number of presentations
Information Science	42,864	12	Linguistics	159,910	27
technology	27,612	4	Gender studies	73,247	11
Statistics	27,612	4	Cultural Studies	52,670	8
Natural resources	17,075	1	Education	46,808	7
Physics	14,851	1	Economics	17,767	3
Chemistry	10,109	1	Management	15,856	4
Medicine	8,382	3	Philosophy	11,393	4
Maths	7,501	1	Communication	9,507	5
			Public Policy	7,475	4
			Regional Studies	5,782	2
			Political Study	4,772	1
			Sociology	4,565	1
Total	156,006	27	Total	409,752	77

Once the program and the word lists for the analysis had been determined, the text files of the transcripts were cleaned up. Following previous studies (e.g., Webb & Rodgers, 2009a, 2009b), speakers' names (e.g., *SU-m, SI*) and nonverbal actions (e.g., *laugh, writing on board*) were removed. Proper nouns and marginal words that appeared in the corpus but were not listed in Nation's (2012) lists of proper nouns (e.g., *Stockhom, Linux*) and marginal words (e.g., *mhmm, whoohoo*) were then added to the relevant lists. After that, the conference presentation corpus was run through RANGE with the 25 BNC/COCA word lists, the list of proper nouns, and the list of marginal words serving as the baseword lists. Similarly, to calculate the percentage of the ASWL word families in the conference presentation corpus, the corpus was analyzed with RANGE with the ASWL acting as the baseword list. Following Author *et al.*'s (XXXXa) approach, the potential coverage achieved by learners with different prior knowledge of general vocabulary from learning the ASWL was the sum of the coverage of (a) the BNC/COCA 1,000 word levels that are relevant to learners' current vocabulary levels, (b) the ASWL words that are beyond their current vocabulary levels, (c) proper nouns, and (d) marginal words.

3. Results

This section presents the results in the order of the research questions. The first research question concerns the vocabulary size needed to achieve 95% and 98% coverage of conference presentations. Table 3 presents the lexical profile of the conference presentations corpus.

Table 3

Lexical profile of conference presentations

BNC/COCA word families (plus proper nouns & marginal words)	Whole corpus	Hard science	Soft science
3,000	96.84 ^a	96.45 ^a	96.99 ^a
5,000	98.22 ^b	97.96	98.31 ^b
6,000	98.49	98.25 ^b	98.57
8,000	98.83	98.67	98.89
10,000	98.98	98.87	99.02

^a The 95% coverage is reached

^b The 98% coverage is reached

The second column of the table shows that the most frequent 3,000 BNC/COCA word families provided 96.84% coverage of conference presentations while the most frequent 5,000 word families provided 98.22% coverage. The next two columns of the table demonstrate the lexical profile of the hard and soft sub-corpora. Both required 3,000 word families to reach 95% coverage. These word families covered 96.45% of the words in the hard science sub-corpus and 96.99% of the words in the soft science sub-corpus, respectively. To reach 98% coverage, hard sciences required a larger vocabulary size (6,000 word families) than soft sciences (5,000 word families).

The second research question asks about the coverage of the ASWL in conference presentations. This list covered 87.52% of the conference presentation corpus, which is consistent with its coverage in each of the two disciplinary sub-corpora: 87.99% (hard sciences) and 87.34% (soft sciences). All ASWL word families occurred in the conference presentations. Appendix 1 lists the 50 most frequent lexical items in each ASWL level that occur in both sub-corpora². However, several word families are common in the conference presentations, but are not ASWL words. Appendix 2 lists the lexical words which are among the most frequent items in the conference presentation corpus but are not included in the ASWL.

The third research question examines the potential coverage that list users having different levels of general vocabulary may achieve if they learn the ASWL. Table 4 presents the potential coverage achieved by users with the support of the ASWL.

Table 4

Potential coverage achieved by users with support of the ASWL (%)

Current vocabulary levels (word families)	Number of ASWL word families beyond users' current vocabulary level	Potential coverage (%)
Fewer than 1000	1,741	94.56
1,000	911	94.88
2,000	455	95.56
3,000	75	96.98

The first row of the table shows the potential coverage by users who have not mastered the most frequent 1,000 BNC/COCA word families. Given their current vocabulary knowledge, they are unlikely to know the ASWL. If these users learn the most frequent 2,000 BNC/COCA word families, they may be able to achieve 92.36% coverage. However, if they learn all 1,741 ASWL word families instead, they will need to learn 259 fewer words families, but can achieve higher coverage of conference presentations (nearly 95%). From a cost-effectiveness perspective, learning the ASWL would seem more useful for these users.

The second row of the table demonstrates the potential coverage achieved by users who have mastered the most frequent 1,000 BNC/COCA word families. Considering their current vocabulary knowledge, these learners could study 1,000 word families from the 2nd 1,000 BNC/COCA word level and achieve about 92% coverage, which may lead to ‘relatively high’ but unstable comprehension rates. However, if they learn the ASWL instead, they can skip Level 1 of the ASWL and focus on learning 911 ASWL word families that are outside the most frequent 1,000 BNC/COCA word families. Such knowledge may allow them to gain potential coverage of approximately 95%, which could lead to ‘relatively high and stable’ comprehension. Therefore, compared to learning the word families from the 2nd 1,000 BNC/COCA word level, learning the ASWL words that are outside their current vocabulary knowledge means that learners would need to learn 89 fewer word families but are likely to achieve a higher level of comprehension.

The third row of the table shows that users having mastered the most frequent 2,000 BNC/COCA word families may need to learn only 455 ASWL word families outside their current vocabulary level because they can skip Levels 1 and 2 of the ASWL. However, they may then be able to recognize more than 95% of the total words in the conference presentation corpus. The potential coverage gained by users with knowledge of the most frequent 3,000 BNC/COCA word families is presented in the last row of Table 4. These users may skip Levels 1, 2 and 3 of the ASWL and only need to learn 75 word families from the ASWL Level 4. This may then allow them to achieve potential coverage of nearly 97%. On the whole, learning 75 to 1,741 ASWL word families may help learners to achieve potential coverage from nearly 95% to 97% of conference presentations. In contrast, if they decide to focus on general vocabulary

instead, learners would need to learn 3,000 word families from the BNC/COCA lists to achieve a similar amount of coverage.

4. Discussion

4.1. What are the lexical demands of academic conference presentations?

Together with proper nouns and marginal words, a vocabulary size of 3,000 word families of general English provides 95% coverage and a vocabulary size of 5,000 word families provides 98% coverage of academic conference presentations. The present study also found that both hard science subjects and soft science subjects require a vocabulary size of 3,000 word families plus proper nouns and marginal words to reach 95% coverage. This vocabulary size is consistent with the size required to reach 95% coverage of the whole corpus. However, to achieve 98% coverage, a larger vocabulary size is likely to be needed for hard science subjects (6,000 word families) than for soft science subjects (5,000 word families). This indicates that perhaps conference presentations in hard science subjects are lexically more demanding than those in soft science subjects. Due to the lack of balance of the sub-corpora, this finding should not be generalized. Yet it is in line with those from earlier studies which showed that larger vocabulary sizes are needed to reach 95% and 98% coverage of lectures and seminars in hard science disciplines (Life and Medical Sciences and Physical Sciences) than in soft science disciplines (Arts and Humanities, Social Sciences) (Author *et al.*, XXXXb), and that lectures, seminars, labs, and tutorials in hard sciences contain a larger number of low frequency words than those in soft sciences (Author, XXXXa). This suggests that from the lexical coverage perspective, understanding spoken academic discourse in hard sciences is likely to be more challenging than in soft sciences.

This study indicates that perhaps conference presentations are as challenging as labs and tutorials, but are less challenging than lectures and seminars in terms of lexical coverage. The number of word families in general English needed to reach 95% and 98% coverage of conference presentations (3,000 and 5,000) was fairly similar to those needed to reach 95% and 98% coverage of labs (3,000 and 7,000) and tutorials (3,000 and 4,000) (Author *et al.*, XXXXc), but smaller than those needed to achieve 95% and 98% coverage of lectures and seminars (4,000 and 8,000) (Author *et al.*, XXXXb)

The similarity in the lexical demand of conference presentations and those of labs and tutorials might be the result of the dialogic nature of these discourse types. Dialogue, which includes the language of all participants of the conversation, is less lexically dense and therefore has a lighter vocabulary load than monologue (Lynch, 2011). An essential component of a conference presentation is the Q & A session, in which presenters interact with the audiences to discuss issues related to the content of their presentations (Rowley-Jolivet & Carter-Thomas, 2005). The purpose of Q & A sessions is similar to the purpose of labs and tutorials, in which students discuss with their tutors and course mates the content and problems related to the lectures (Author et al, XXXXa).

The fact that conference presentations had lighter lexical demands than academic lectures may be due to the nature of these discourse types. Conference presenters are given less time and space to present the content than academic lecturers, and people attending a conference presentation session usually have some background knowledge on the topic presented in that session (Rowley-Jolivet & Carter-Thomas, 2005). Therefore, instead of reviewing specialized concepts and theories in detail, conference presenters often briefly mention the terminology and focus more on sharing the key findings and highlighting the contributions of their studies (Rowley-Jolivet & Carter-Thomas, 2005). This may lead to the lower percentage of low frequency words in academic conference presentations compared to that in academic lectures, which explains why conference presentations have lighter lexical demands than academic lectures.

4.2. Is the ASWL a useful tool to improve comprehension of academic conference presentations?

The ASWL itself covered more than 87% of the words in the academic conference presentation corpus as well as its two disciplinary sub-corpora. This coverage figure is lower than the coverage of the ASWL in academic lectures, seminars, tutorials, and labs reported by Author *et al.* (XXXXa) (around 90%). The difference in coverage may be because of the nature of the ASWL. This list was originally developed to enhance comprehension of class events (lectures, seminars, tutorials, and labs) rather than non-class events such as conference presentations. Nevertheless, the findings of the current study indicate that the ASWL is still a useful resource for EAP programs to support comprehension of conference presentations for two reasons.

First, the ASWL can benefit learners irrespective of their vocabulary levels. It is a shortcut for those with insufficient knowledge of the most frequent 3,000 BNC/COCA word families to achieve adequate comprehension of conference presentations. If learners with insufficient knowledge of the most frequent 1,000 word families of general vocabulary learn the most frequent 2,000 BNC/COCA word families, they could potentially recognize about 92% of the running words in conference presentations, which may result in ‘relatively high but unstable’ comprehension. However, if these learners learn 1,741 word families from the ASWL, they would need to study 259 fewer word families, but may be able to achieve nearly 95% coverage, which is likely to lead to ‘relatively high and stable’ comprehension.

Likewise, learners having mastered the most frequent 1,000 word families of general vocabulary may need to learn only 911 ASWL word families that are beyond their current vocabulary level. Yet this may allow them to reach approximately 95% coverage of conference presentations. In contrast, learning 1,000 word families from the 2nd 1,000 BNC/COCA word level only helps them to achieve around 92% coverage. Similarly, instead of learning 1,000 word families from the 3rd 1000 BNC/COCA word level, learning 455 ASWL word families outside their current vocabulary knowledge may still enable students with the vocabulary knowledge of the most frequent 2,000 word families to recognize more than 95% of the running words in conference presentations. These findings are meaningful given that research has reported that a reasonable proportion of L2 learners have low language proficiency in their second languages (e.g., Akbarian, 2010; Cheng, 2016).

As for learners with sufficient knowledge of the most frequent 3,000 word families, given their current vocabulary knowledge, they may be able to recognize 95% of the running words of conference presentations. However, this does not mean that the ASWL is not useful for them. The greater the coverage, the better comprehension is (Schmitt et al., 2011). Learning only 75 word families from the ASWL that are beyond their current vocabulary knowledge (e.g., *triangle, arbitrary, optimize, inverse*) may allow these learners to approach closer to the 98% coverage, which indicates ‘very high’ listening comprehension.

A second reason why the ASWL is a useful resource is that it can potentially have wide applications. The list is useful for EAP programs which consist of learners from both hard sciences and soft sciences. It consistently covered more than 87% of the hard and soft subject

sub-corpora. Moreover, EAP learners only need to learn items from a single list but may be able to deal with different kinds of speech events which are important for their subsequent academic study and career (lectures, seminars, labs, and conference presentations).

The findings of the present study should be interpreted with caution. First, the corpus of conference presentations was unbalanced due to the challenge of collecting spoken data, especially conference presentations. Second, while lexical coverage is a key factor, other factors may influence the comprehension of academic conference presentations such as background knowledge, listening strategies (Bonk, 2000; Wallace, 2020), speakers' accents (Goh, 2000), speech rate, degree of concentration, visual aids, and interaction (Flowerdew & Miller, 1992; Forey & Feng, 2016). Third, this study only examines vocabulary in conference presentations from the perspective of single words. Future research which is based on a more balanced corpus and examines both single word and multiword units may provide further insights into the nature of vocabulary in conference presentations. Fourth, the present study highlights the value of the ASWL for EAP learners based on corpus analyses. Research investigating the implementation of the list in the real classroom would be useful. Fifth, this study only examined the vocabulary needed for comprehending conference presentations. Studies examining the vocabulary needed for delivering conference presentations effectively would be valuable.

5. Pedagogical implications

This section provides some suggestions on how EAP teachers can make use of the ASWL to help their learners' deal with the lexical demands of conference presentations. To begin with, teachers should estimate the challenge of understanding conference presentations for their learners so that they can provide relevant support. They could use Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT) to measure learners' knowledge of the most frequent 5,000 word families. The online and PDF versions of the test as well as the answer key are available on Stuart Webb's website: <https://www.edu.uwo.ca/faculty-profiles/stuart-webb.html>. The present study found that while the lexical demands of conference presentations may vary across disciplines and contexts, the most frequent 3,000 word families are necessary for adequate comprehension of conference presentations. Therefore, if the UVLT scores show that learners have mastered at least the most frequent 3,000 word families, they are likely to comprehend conference presentations. In this

case, teachers could advise learners to learn the 75 word families from Level 4 of the ASWL so that they could reach a higher level of listening comprehension. If the UVLT scores show that the learners are yet to master the most frequent 3,000 word families, it means that they are likely to have great difficulty in comprehending conference presentations. In this case, depending on learners' current vocabulary level, teachers can use Author *et al.*'s (XXXXa) model as a guide to set the learning goals for their learners. As the model has been discussed in detail in other publications (e.g., Author, XXXXb; Author *et al.* XXXXa), this section only describes it briefly. According to this model, if learners want to focus on items that occur frequently in academic spoken English, they can learn words from the ASWL level that are relevant to their current knowledge of general vocabulary. Such knowledge would enable them to recognize a large proportion of words in conference presentations as well as other kinds of academic spoken English (e.g., lectures, seminars, labs, and tutorials). However, if learners want to broaden their general vocabulary repertoire first, they can focus on the words in the subsequent BNC/COCA word level and later they can focus on the ASWL level that is relevant to their knowledge of general vocabulary at that stage. In addition to the ASWL words, teachers could also focus learners' attention to the items listed in Appendix 2, which are frequent in the conference presentation corpus but are not in the ASWL.³

Once the learning goals have been set, teachers could design materials and activities to help learners meet and use the ASWL word families repeatedly by following Nation's (2007) Four Strands: meaning-focused input, meaning-focused output, fluency development, and language-focused learning. Each learner can use some search engines (e.g., Google, Youtube) to find a conference presentation in their specific area. For example, a learner majoring in Applied Linguistics/TESOL may watch Nesi and Alsop's (2015) presentation, 'Variation and commonality in EFL and English L1 engineering lectures from around the world' at the BALEAP 2015 conference via this link <https://www.youtube.com/watch?v=DVMrYs88lcI>. To have some idea of the lexical profile of the presentation, they can use the *Vocabulary Profiler* function of Lextutor <https://www.lexutor.ca/vp/> to analyze the transcript of the presentation. As shown in Appendix 3, knowledge of the most frequent 3,000 word families is needed to reach 95% coverage of this presentation, which is in line with the findings of the present study. After watching this presentation, the learner can search for readings related to the content of the presentation (e.g., the features of academic lectures in different contexts). Based on his/her

understanding of the content of the presentation and the readings, s/he can then write a reflection and orally present it to his/her classmates. These linked skills activities would create good conditions for learners to reencounter and use the ASWL word families in a meaningful way (Webb & Nation, 2017).

Apart from these meaning-focused activities, teachers can organize some language-focused learning activities in which they explicitly draw learners' attention to the linguistic features of the ASWL words. For example, teachers can use the wordlist function in Antconc with the ASWL as the baseword list to identify the ASWL word families that occurred in the script of a conference presentation and design activities for learners to notice and learn the ASWL word families. To illustrate some possible language-focused activities that teachers can do, let us take Nesi and Alsop's (2015) presentation and the word *know*, *model*, and *differentiate*, which are in the top 50 most frequent lexical items in each ASWL level, as examples.

For the ASWL level 1-3 words, which are also common in general conversation, teachers can elicit from the students their knowledge of the meaning and use of these words (e.g., *know*, *model*). Then, they can provide students with a set of concordance lines of these words in the conference presentation transcript (see Appendix 4 for *know* and Appendix 5 for *model*) and ask them to identify the typical patterns and meaning of these words in the conference presentation. During this process, teachers can facilitate a follow-up discussion to help learners realize that apart from using *I/we/you/she/know(s)* as a way to present new information as in written English, *you know* is more likely to be used as a discourse marker to engage the listeners, and that while in general English, *model* can be used as a noun to refer to a fashion model, in the examined conference presentation, this word is used with a slightly different meaning and can also be used as a verb.

For the ASWL Level 4 words (e.g., *differentiate*), which are uncommon in general conversation, teachers could refer back to the texts surrounding these words (see Appendix 6) and ask the learner to guess the meaning from the context. If learners are not able to do so, teachers can directly teach the words. Together these language-focused learning activities would help learners to broaden their knowledge of the ASWL words such as technical meanings, collocations, spoken-written forms, and parts of speech.

Apart from setting vocabulary learning goals and organizing activities to achieve them, it is important for teachers to raise learners' awareness of the value of the ASWL in supporting their comprehension of conference presentations. This could be done by referring to the findings of the current study and the information in Tables 3 and 4 so that learners are aware of the lexical demands of conference presentations and how the ASWL can help them to meet this demand.

6. Conclusion

The present study is the first to investigate the lexical demands of conference presentations. Based on the analysis of vocabulary in the largest corpus of conference presentations ever created, it found that to reach 95% coverage, learners need a vocabulary size of the most frequent 3,000 word families. The ASWL itself provided more than 87% coverage. Learning from 75 to 1,741 ASWL word families that are beyond their current vocabulary levels may allow users with different levels of general vocabulary to reach coverage from nearly 95% to 97% of conference presentations. The coverage of the ASWL in the two disciplinary sub-corpora is very consistent, which indicates that the list would be a useful resource for both hard science and soft science students. Together the findings suggest that the ASWL is a useful instrument for EAP programs for research students and scholars with different prior vocabulary knowledge and from both hard sciences and soft sciences.

Acknowledgements

My thanks to Cambridge University Press and the researchers in the Michigan Corpus of Academic Spoken English project, the John Swales Conference Corpus project, the English as a Lingua Franca in Academic Setting corpus project, and the Vienna-Oxford International Corpus of English project for allowing me to use their materials to create the academic conference presentation corpus.

Notes

¹ A valid corpus-based word list should provide similar coverage in the corpus from which it was developed and in an independent corpus of similar genre (Nation & Webb, 2011).

² Items selected for each level should (a) be lexical words in the ASWL, (b) appear in both the hard and soft sub-corpora of the conference presentation corpus, (c) have Juilland D dispersion of at least 0.6, and (d) rank 1st to 50th in terms of frequency in the conference presentation corpus.

³ Further concordance analysis showed that several of these words are related to the practicability of conferences (*coffee, ice, chair, guest, seminar, poster, lunch, food, tonight, evening, February*). These items are not closely related to the subject-specific content of the conference presentation and may have less value for learning than the other items.

References

- Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38(3), 391–401.
- Biglan, A. (1973a). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, 57(3), 204–213.
- Biglan, A. (1973b). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57(3), 195–203.
- Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14, 14–31.
- Cheng, A. (2016). EAP at the tertiary level in China. In *The Routledge Handbook of English for Academic Purposes* (pp. 97–108). Routledge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A. (2011). The Academic Word List 10 year on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–362.
- Coxhead, A. (2016). Reflecting on Coxhead (2000), ‘a New Academic Word List’. *TESOL Quarterly*, 50(1), 181–185.
- Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 97–110). Routledge.
- Feak, C. (2013). ESP and Speaking. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 35–54). Wiley-Blackwell.
- 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.
- Forey, G., & Feng, D. (2016). Interpersonal meaning and audience engagement in academic presentations: A multimodal discourse analysis perspective. In *The Routledge Handbook of English for Academic Purposes* (pp. 416–430). Routledge.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75.
- Goh, C. C. M. (2013). ESP and Listening. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 35–54). Blackwell.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range: A program for the analysis of vocabulary in texts*. Retrieved from <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Hincks, R. (2003). Pronouncing the Academic Word List: Features of L2 student oral presentations. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences*.
http://www.speech.kth.se/ctt/publications/papers03/icphs03_1545.pdf
- Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405–411.
- Lynch, T. (2012). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10, 79–80.
- Morell, T. (2015). International conference paper presentations: A multimodal analysis to determine effectiveness. *English for Specific Purposes*, 37, 137–150.
- 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 <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. 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 University Press.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Nesi, H., & Alsop, S. (2015). *Variation and commonality in EFL and English L1 engineering lectures from around the world*. BALEAP, Leicester.
<https://www.youtube.com/watch?v=DVMrYs88lcI>
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Continuum.
- Pérez-Llantada, C., Plo, R., & Ferguson, G. R. (2011). “You don’t say what you know, only what you can”: The perceptions and practices of senior Spanish academics regarding research dissemination in English. *English for Specific Purposes*, 30, 18–30.
- Reershemius, G. (2012). Research cultures and the pragmatic functions of humor in academic research presentations: A corpus-assisted analysis. *Journal of Pragmatics*, 44, 863–875.
- Rowley-Jolivet, E., & Carter-Thomas, S. (2005). The rhetoric of conference presentation introductions: Context, argument and interaction. *International Journal of Applied Linguistics*, 15(1), 45–70.
- 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.
- Shalom, C. (1993). Established and evolving spoken research process genres: Plenary lecture and poster session discussion at academic conferences. *English for Specific Purposes*, 12, 37–50.
- Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In K. 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.

- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56–89.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258.
- Wallace, M. (2020). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*.
<https://doi.org/10.1111/lang.12424>
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Webb, S., & Rodgers, M. (2009a). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407–427.
- Webb, S., & Rodgers, M. (2009b). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366.
- Webb, S., Sasao, Y., & Balance, O. (2017). The Updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, 168(1), 34–70.