A QUESTION OF GENERATED WORD LISTS

Ross Sampson
Kanda University of International Studies

ABSTRACT

As Saville-Troike (1984 as cited in Jordan, 1997, p.149) states, "vocabulary knowledge is the single most important area of second language competence regarding academic achievement". The words used to convey meaning are so important, in order to do well academically and thrive in your disciplinary discourse community, which means knowing which words to use is vital. This article examines learning strategies used to utilise the Academic Word List (Coxhead, 2000) within an English for Academic Purposes (EAP) coursebook. It concludes that a general list of useful academic words can benefit students, however it suggests that more work needs to be done in order to cater to specific student needs and additionally to generate more spoken corpuses. This paper will conclude with the study's relevance to the academic writing context in KUIS.

INTRODUCTION

In many university contexts it is difficult to adequately supply students with the necessary vocabulary that they need. This is why investigating a general academic provision of vocabulary and its feasibility is of value to English for Academic Purposes (EAP) . The aim of this study is to put a spotlight on the Academic Word List (AWL) as an optimal general academic list of vocabulary for the university context. This study also hopes to provide the reader with a clear understanding of the work that has been done related to academic word list corpora and enable teachers to assess what the vocabulary requirements of their own learners may be. In the context of universities in Japan, and in particular KUIS, this study may prove useful for teachers, especially teachers whose expertise may not lie in vocabulary. At KUIS, many courses, particularly those with titles related to academia, have clear vocabulary components. Many teachers are aware of the need for learners of English to build on their breadth and depth of vocabulary knowledge. Therefore, after reading this article they may consider the question of discipline and context specific vocabulary needs versus general vocabulary provision.

Insufficient vocabulary knowledge has been connected to a gap in academic achievement, primarily by English language learners (Chall, 1996; Hart & Risely, 1995; Neuman, 2008; Hierbert & Lubliner, 2008; Risley, 1995; Townsend, Filippini, Collins, & Biancarosa, 2012 as cited in Gardner & Davies, 2014, p. 305). A lack of vocabulary can have a significant effect on status within a university and has been identified as an obstacle to success (Corson, 1997; Snow & Kim, 2007). Corson (1997) argues that a good grasp of academic vocabulary is evidence of being in control of academic meaning systems which lead to success.

LITERATURE REVIEW

Academic vocabulary is words which appear with a much greater frequency in academic texts than other types of texts (Townsend & Kiernan, 2015). This can be difficult to fully master (even for native speakers) for a number of reasons; it is infrequent, it conveys abstract ideas and it is largely Graeco-latin (Nation, 2001). It contains morphologically complex words, many parts of speech are nominalized and it is information dense (Biber, 2006). However, all these aspects are interrelated and together produce academic communicative functions.

Research into corpus linguistics has brought to light the fact that vocabulary differs across registers. A register such as academic texts must be properly represented to account for disciplines on opposite ends of the academic spectrum (hard sciences – applied subjects). Representation is crucial for a list claiming to be 'general academic'. Issues of the number, length and types of texts to include are debated (Biber, 1993; Sinclair, 1991 as cited in Coxhead, 2000, p. 215). Sinclair (1991) warns that a corpus which "does not reflect the size and shape of the documents from which it is drawn, is in danger of being seen as a collection of fragments where only small-scale patterns are accessible" (p. 19).

A number of attempts have been made to produce a list of the most useful general academic words. In 1984, Xue and Nation (Coxhead, 2000; Gardner & Davies, 2014; Nation & Hwang, 1995) created the University Word List (UWL). This was a combination of four previous lists (by Campion & Elley, 1971; Ghadessy, 1979; Praninskas, 1972), these lists were compiled all before the internet, meaning they would have been very time consuming. This list was used by many teachers and course designers; however, its selection principals (similar to previous lists) meant it failed at representing a comprehensive academic list.

Nowadays, the most widely known and used academic list is one compiled by Avril Coxhead (2000) which was rigorously validated. Coxhead created the AWL with the intention of providing a tool to guide teaching, learning and materials design. It was done by investigating the frequency and range of words outside the first 2000 most frequent words of English. The total corpus made up three and a half million running words of written academic text. The AWL focused on word families (a headword plus all inflected and derived forms Nation, 2001) as this solved the question of what constituted a word (Coxhead, 2000). Coxhead (2015) points out "selecting items for teaching is a major area of concern because of the cost-benefit equation, learners and teachers need to get the best return for effort" (p. 641). The criteria for word selection looked at specialised occurrences, which are words found outside the most frequent two thousand. Range was a focus, this meant a member of a word family had to occur at least ten times in each of the four main sections and in fifteen or more of the subject areas. Frequency was also a focus, this meant a member of a word family had to occur at least a hundred times in the corpus (Coxhead, 2000, p. 221). Range was prioritised over frequency, as word count favouring frequency would have been biased by longer texts. Coxhead found the difference in coverage across the four disciplines to favour Commerce students (2000). She stated that her AWL words occurred in a wide range of subject areas with one hundred and seventy-two of the five-hundred and seventy word families occurring in all twenty-eight subject areas (Coxhead, 2000, p. 222), and ninety-four percent of the word families occurred in twenty or more subject areas. She claimed that the first 2000 words in the General Service List (GSL) (West, 1953) combined with the AWL make up about eighty-six percent of the academic

corpus, suggesting that pursuing the AWL can increase learners' understanding of texts significantly. Coxhead used a second corpus to test the AWL's coverage rate and found all word families accounted for, with an 8.5 percent coverage (Coxhead, 2000). She also found that by testing the AWL in a non-academic corpus it could be concluded that the majority of word families (410 clearly academic) in the AWL could be understood as academic (Coxhead, 2000, p. 224). She finally compared her AWL with the previously used and highly regarded UWL, and claimed that despite the AWL being smaller, it had a higher and thus better coverage of academic texts as well as a wider subject area range (Coxhead, 2000, p. 226).

Despite the AWL being held to such a high regard, it naturally has critics. Durrant (2014) states that researchers have suggested there may not be a generic academic vocabulary of use for all EAP learners. This is because words are not evenly dispersed and vary in use and across disciplines. Hyland and Tse (2007) have criticised Coxhead's list as they believe corpus analysis should inform vocabulary teaching in EAP but should be more geared towards specific needs rather than general. The specific vs. general vocabulary debate continues to be a point of contention among researchers in this area. Hyland and Tse describe Coxhead's AWL as "opportunistic", as it does not fairly represent each discipline. They say – "the fact that all disciplines shape words for their own uses seriously undermines attempts to describe a core academic vocabulary" (2007, p. 240). Already holes can be seen in the AWL that make it appear not as robust as once thought.

Hyland and Tse (2007) conducted a study which created a corpus containing; long and short texts, peer-reviewed articles, pedagogic texts and student writing from disciplines, making up three sub-corpora of: engineering (mechanical and electronic), sciences (biology, physics, and computer science) and social sciences (sociology, business and applied linguistics). They found 534 AWL word families (93.7%) had irregular distributions across these sub-corpora and 40% (227) of these had at least 60% of all instances in just one sub-corpus. Only 36 word families were found to be evenly distributed, meaning only a much smaller general academic word list could be provided. They continued to show from their study that the unevenness also occurs within each sub-corpus, stating that only one-word family occurred equally across the three science disciplines, seven evenly in social sciences and forty-seven in the engineering disciplines. They suggested items have a limited range (semantic variation) across sub-corpora, which means a more complex picture of language in use. They also cautioned, as most words are polysemous, we must be cautious about generalising word families where "meaning could differ across each inflected and derived word form" (Oakley, 2003, as cited in Hyland & Tse, 2007, p. 243). Schmitt and Zimmerman (2002), in a study on derived word forms, found it was uncommon for learners to know all word forms, which means knowing one word in a word family does not necessarily mean knowing the others. Hyland and Tse (2007) showed that one of the most frequent AWL words in the corpus, *analyse*, is used differently across fields (e.g. as a noun in social science but as *analytical* in engineering). They say the AWL would be most useful to computer science students (16% coverage by the list) and least useful to biology students (6.2% coverage).

Gardner and Davies (2014) also criticise the AWL, but do not think general academic lists should be dismissed. They proposed a new list of core academic vocabulary based on lemmas as opposed to word families. Their goal was to create a core list excluding general-high frequency and discipline specific words (2014). They used a corpus thirty-five

times larger than the AWL. Although all texts used were published in the USA and it too only contained written English. They claimed that basing it around lemmas meant teachers and learners could focus on the most frequent lemmas in a word family (2014), helping direct academic needs more. They called this list the Academic Vocabulary List (AVL). It is unclear if any coursebooks have utilised Gardner and Davies' new AVL or not. There definitely appears to be improvements from the AWL, and in theory, the larger the corpus the better, but as Hoey (as cited in Lewis, 2000, p. 227) cautions, "word list[s] can be dangerous", as "none show any awareness of collocation".

Studies have shown (Campion & Elley, 1971; Hwang, 1989 as cited in Nation 2001, p. 189) that it is possible to provide academic vocabulary for a range of academic disciplines. Although as Nation and Chung (2004) comment, "technicalness" is a functional aspect of a word, which means its use determines how technical it is. Therefore, words could (technically) be both general and technical academic words depending on the context. For example, students in a business studies course might not know the word *firm* for a company or *plant* for a factory (Alexander, Argent, & Spencer, 2008). To add to all this, as technical words are used more often, they become less technical. Based on these critiques, it can be seen that there are issues with the AWL. However, I argue that despite these issues, the AWL is a robust offering of academic vocabulary and is still able to provide the best general academic words covering a wide range of subjects. This is why I am in favour of Coxhead's list being used in university settings and for academic writing contexts generally.

METHODOLOGY

The context for this investigation of vocabulary is a university class of mixed disciplinary backgrounds. In this context it is difficult to supply individual students with the vocabulary and ways of creating discourse pertinent to each of their fields of study. The findings related to this study may be applicable to the KUIS context.

Nation (2001) offers three vocabulary categories;

1. High frequency words – This includes the two thousand most widely used words (eighty percent coverage of most texts).

2. Academic or sub-technical (frequent in academic writing across disciplines and genres).

3. Technical vocabulary (low frequency and difficult depending on discipline).

Of these, it is academic vocabulary which is the focus here. Coxhead (2000) suggested her AWL be used to create teaching materials and reach EAP vocabulary needs and goals, and hoped authors would write books based on the AWL. This is exactly what Diane and Norbert Schmitt did in *Focus on Vocabulary 2: Mastering the Academic Word List*, which is the textbook used as the material of investigation. I chose the textbook as it purports to utilise the AWL through the vocabulary activities within it. Townsend and Kiernan (2015) make the interesting point that "rarely are connections made between corpus linguistics and classroom practice". It could be said that corpus data can be of great application for EAP teaching and this book seems to have academic vocabulary at heart. What was hoped to be understood through this investigation was:

1) To what extent does this book exemplify the AWL in the way in which it claims?

2) To what extent is the AWL the optimal academic vocabulary provision?

The book has seven units with four chapters in each. In the first three chapters of each unit 72 words are introduced and in the fourth, strategies are exemplified with consolidation and recycling of vocabulary. Research has shown multiple opportunities in multiple contexts can help acquisition (Blachowitz & Fisher 2000), so recycling vocabulary is an essential part of the learning process. As Resnick (1989, as cited in McCarthy & Schmitt 1997) states "when instructors model explicit methods of vocabulary acquisition which require deep processing and re-encountering of words, students not only learn words but experience 'cognitive apprenticeship'" (p. 255). This means students learn that vocabulary acquisition is a task requiring their active participation.

The *strategy practice* at the end of each chapter contains a *using your dictionary* subsection, a *strategy* subsection and a *word knowledge* subsection. These sections are designed to explicitly inform learners about how to become autonomous in their vocabulary learning. Adult learners especially, should not be reliant on the teacher and should recognise their active participation and individual pursuit of learning is vital. A responsible learner is one "who accepts the idea that their own efforts are crucial to progress in learning and behaves accordingly" (Scharle & Szabo, 2000, p. 3). Understanding learning strategies can be key to promoting autonomy in learners. Nation (2001, p. 217) says it can be difficult to define a strategy, but it should:

1. Involve choice (several strategies to choose from).

2. Be complex (several steps to learn).

3. Require knowledge and benefit from training.

4. Increase the efficiency of vocabulary learning and use.

These definitions by Nation are utilised in this study as a checklist in the examination of the chapter 12 section on *strategy practice*. The words listed in Appendix A are all previously introduced in the three chapters of unit three (in authentic texts and exercises) before my materials (chapter 12).

ANALYSIS

Using your dictionary subsection
This section is about *grammar information*. A study by Harvey and Yuill (1997) found that learners did not fully utilise dictionaries grammatical coding schemes, even though Nation (2001) says dictionaries can help learners with vocabulary. In this section, explanations are first provided to help learners recognise how understanding grammar information can benefit their vocabulary knowledge. The first example *potential* indicates to the learner that this is not a countable noun, the second – common collocations and the third – that possessive pronouns precede this word. All three attempt to convey how aspects of grammar affect words. If learners are able to use dictionary entries to their full potential, it will greatly aid their vocabulary knowledge development. After explicitly detailing how grammar information can be applied, this section then gives three more entries (AWL words previously introduced) and a task in which learners must correct grammatical

mistakes (made by actual learners) by using the dictionary information provided. Using the dictionary effectively requires learner's knowledge as it is complex and not obvious how to optimally use a dictionary. The dictionary is a much used and valuable resource for language learners; however, it should be used effectively and not relied upon.

Strategy subsection

This section is about *guessing from context*. This technique is widely understood as one of the most important learner strategies for tackling unknown words. "Studies have shown a large proportion of words (at least eighty percent) can be guessed" (Liu & Nation, 1985, p. 35). Guessing from context relates to incidental learning but needs to be explicitly explained, as learners may have an over reliance on their dictionary and see using it as the first port of call when encountering challenging vocabulary. It appears the steps proposed for learners in this section are from Nation and Clarke (1980, as cited in Nation 2001, p. 257). These are language clues aimed at developing fluency so as not to overly disrupt the flow of reading. There is a step-by-step progression from a narrow focus (e.g. parts of speech) to a broader focus (e.g. to guess the meaning using tools introduced in each step). This strategy is complex, there are clear benefits from training, and it does aid vocabulary learning and use (although not a flawless technique as wrong guesses can be made).

Word knowledge subsection

This section, titled *frequency of occurrence*, deals with generated word lists and frequency. A word frequency table is provided showing word frequency for spoken and written English (according to the *Longman Advanced American Dictionary*). The task is a gap-fill in which learners must use a given academic word (all from the AWL and previously introduced) to replace a general English word. This task provides a context and encourages learners to consider how general English words can be replaced to make them more academic. This involves choice, requires learners' knowledge of the words and increases efficiency and vocabulary use by making the distinction between general and academic.

DISCUSSION

The AWL should definitely be considered as an academic vocabulary provision for university students. These materials are good for encouraging learners to recognise how a dictionary can be utilised, how to be autonomous in vocabulary learning and how to select academic words. If learners can *notice* (Schmidt, 1990) aspects of language and compare with their previous encounters of vocabulary in context, they should be able to benefit and develop their academic vocabulary knowledge. The strategy sections appear to meet Nations' (2001) criteria for what a strategy is, and thus exemplify the AWL well. Therefore, it is a very worthy academic list that can truly serve learners in university contexts. In the KUIS context, many learners are required to produce research papers and thus sometimes struggle to produce the correct academic language. The AWL can help them get a better grasp of understanding and using academic vocabulary.

It may indeed be impossible to provide a vocabulary list able to cater to every academic learners' specific needs, but providing learners with a general list along with suitable activities and strategies for noticing and encouraging autonomy can go a long way to aiding English language learners transition into an academic environment. Furthermore, the word knowledge section mentions spoken English, this is something the AWL does not account for. More work needs to be done to gather real spoken academic texts despite the

obstacles of cost, time and resources. Including spoken academic data is crucial to accurately represent all vocabulary produced in academic settings.

CONCLUSION

Communicative purpose and genre determine what words are used, why they are used and how they are used. Learners need to be directed to understand this, which will help them to analyse and think about the vocabulary they encounter to work out meaning and function. In KUIS, many students take academically focused courses and additionally also take courses in preparation for reputable English proficiency tests such as TOEIC or TOEFL. Being prepared as a teacher for the potential vocabulary needs of your students is something essential in providing the students with the relevant knowledge to help build their vocabulary for a particular course and its content. Being aware as a student and being autonomous in vocabulary learning is also essential. Students should be actively making attempts to deepen their understanding of a wide variety of vocabulary relevant to their needs and taking it upon themselves to find resources in order to strengthen and extend their knowledge. At KUIS many teachers are prepared for this and give explicit attention to the learning of vocabulary in class. Teachers and Learning Advisors in KUIS also guide students to vocabulary strategies and resources that they can use for themselves in an effort to find what works for them. Vocabulary is sometimes an overlooked aspect of language learning, however both educators and learners should be aware of what is needed in their context and how to acquire it best.

REFERENCES

Alexander, O., Argent, S., & Spencer, J. (2008). *EAP essentials: A teacher's guide to principles and practice*. Reading, UK: Garnet Publishing, Ltd.

Biber, D. (2006). Stance in written and spoken university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

Blachowicz, C. L. Z., & Fisher, P. (2000). Vocabulary Instruction. In M. L. Kamil, P. B. Mosethal, P. D. Pearson, & R. Barr (Eds.) *Handbook of reading research* (Vol 3, pp. 255-91). Mahwah, NJ: Erlbaum.

Chung, T.M. & Nation, P. (2004). Identifying technical vocabulary. *System, 32*, (2), 251-263.

Clarke, D.F. & Nation, I.S.P. (1980). Guessing the meanings of words from context: Strategy and techniques. *System, 8*(3), 211-220.

Corson, D. (1997). The learning and use of academic words. *Language Learning, 47*(4), 671-718.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A. (2015). Vocabulary research and pedagogy: Introduction to the special issue. *Language Teaching Research, 19*(6), 641-644.

Coxhead, A. (2016). Reflecting on Coxhead (2000), A New Academic Word List. *TESOL Quarterly, 50*(1), 181-185.

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics,* 35(3), 328-356.

Gardner, D. & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305-327.

Harvey, K., & Yuill, D. (1997). A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. *Applied Linguistics, 18*(3), 253–278.

Hyland, K. (2006). *English for academic purposes: An advanced resource book.* New York: Routledge.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235-253.

Jordan R. (1997). *English for academic purposes: A guide and resource book for teachers.* London: Cambridge University Press.

Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach.* Hove, UK: Language Teaching Publications.

Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*(1), 33–42.

Ming-Tzu, K., & Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics, 25*(3), 291-314.

Nagy, W., Townsend, D., Lesaux, N., & Schmitt, N. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly. 47*(1), 91-108.

Nation P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.

Nation P. (2001*). Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, P. & Hwang. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System, 23*(1), 35-41.

Scharle A., & Szabo A. (2000). *Learner Autonomy*. London: Oxford University Press.

Schmitt N. (2000) *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. & McCarthy, M. (1997*). Vocabulary description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.

Schmitt, N. & Zimmerman, C.B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145-171.

Sinclair J. (1991). *Corpus, concordance, collocation*. London: Oxford University Press.

Snow, C. E., & Kim, Y.S. (2007). *Large problem spaces: The challenge of vocabulary for English language learners.* In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension.* (pp. 123-139). New York: Guilford Press.

Townsend, D. Filipini, D. Collins, P. & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal, 112*(3), 497-518.

Townsend, D., & Kiernan, D. (2015). Selecting academic vocabulary words worth learning. *The Reading Teacher, 69*(1), 113-118.

West, M. (1953). *A General Service List of English words*. London: Longman, Green & Company.

Appendix A: AWL words

Words recycled in chapter 12, the chapter within the same unit which they were introduced and the sub-list of the AWL; e.g. *potential* and its family members are in sub-list 2. Sub-list 1 = most frequent words in the list and 10 = least frequent.

Words in chapter 12	Chapter previously introduced in the textbook	Sub-list of AWL
Potential	10	2
Indication	10	1
Seek	9	2
Undergo	11	10
Facilitate	9	5
Preceding	11	6
Approximately	11	4
Crucial	10	8
Sufficient	11	3
Display	10	6
Authority	10	1
Period	11	1
Assign	9	6
Contract	11	1
Enhance	9	6
Exhibition	10	8
Restrict	9	2
Conclusion	10	2
Corporate	10	3
Exclusion	9	3
Innovation	10	7
Transmit	11	7