ACHIEVING LEXICAL THRESHOLD FOR SCIENCE UNDERGRADUATES LEARNING EAP IN THE ESL CONTEXT

F. S. Izzadeen*

English Language Teaching Unit, University of Peradeniya

INTRODUCTION

The increasing number of students reading for a degree in their second language (L2) has increased the focus on English for Academic purposes in the L2. Although vocabulary had been relegated to the back benches, a surge in vocabulary research has re-established the importance of vocabulary to achieving language competence (Laufer & Nation, 2011). As studies reveal that even at undergraduate level lack of vocabulary creates the most difficulty. This is an issue that needs to be addressed. While academics are in agreement of the need of a minimal coverage of 95 % for text comprehension (Laufer & Ravenhorst-Kalovski, 2010; Nation & Hu, 2000) there are differing views as to how this is best achieved. In the case of academic language the General Service List (GSL) created by West (1953) together with Coxhead's Academic Word List (AWL) (Coxhead, 2000) has been said to provide a coverage of around 90 %. However studies indicate that this does not hold true in the science based disciplines where the coverage by GSL and AWL has been shown to be much less (Nassaji & Vallipouri, 2013; Chen & Ge, 2007). It is also argued that the concept of an academic vocabulary common to all disciplines is not valid (Hyland & Tse, 2007) and attention needs to be paid to the academic vocabulary of each discipline or domain. In order to investigate this claim and also to investigate what vocabulary science undergraduates need to focus on in order to achieve the lexical threshold, a 252,400 word corpus of language encountered by science undergraduates in the discipline of Chemistry was compiled and analyzed. A list of Domain Specific Academic Words in Chemistry (DAWLIC) which would help in targeted EAP courses for science undergraduates was compiled from this corpus.

METHODOLOGY

Compilation of the corpus

Samples for the corpus were selected on the basis of being written academic language of different text types and communicative functions, which undergraduate students encounter in the discipline of Chemistry. Material used at the University of Peradeniya within a time frame of ten years was selected.

The corpus was compiled following principles in corpus design. Special attention was paid to the issues of sample size and type in order to avoid bias and achieve balance and representativeness.

Only part of the material was available in soft copy format. Therefore a majority of the material had to be transformed into a computer friendly format. This was achieved by two methods. Legibly printed material was scanned and converted to .txt format using optical recognition software. A large part of the material was typed via the keyboard. In both these methods material was proofread to ensure accuracy.

The material was divided into four sub-corpora according to the sub-disciplines in chemistry they were from.

Analysis of the corpus

The corpus was analyzed using the corpus analysis software RANGE which can be downloaded at http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx . This software can be used to compare the vocabulary of up to 32 texts at a time. It gives information about the

Corresponding author: Email - **sherzana.i@gmail.com*

frequency of words as well as how many of the texts being compared it occurs in. Using RANGE it is also possible to find out what words from the first and second thousand words in the GSL and Coxhead's AWL are present in a text or corpus being analyzed.

Creation of DAWLIC and the technical word list

In order to create a List of academic words specific to the domain of Chemistry, words belonging to the GSL and AWL were excluded by making use of the data from the analysis using RANGE. Words were selected for inclusion in DAWLIC on the criteria of at least 4 occurrences in the corpus and a minimum range of appearing in at least two of the sub-corpora. Abbreviations, proper nouns and chemical formulae were not included. Technical vocabulary was identified using a rating scale developed on the lines of Nation & Chung's (2003) work. A separate list was compiled of the technical vocabulary thus identified.

RESULTS AND DISCUSSION

Vocabulary coverage of corpus by different word lists

The coverage of the compiled corpus, by the GSL, AWL and other words which do not belong to either category, is presented in table 1. As seen in the table, the first and second thousand GSL words respectively account for 64.71 and 5.87 % of the vocabulary. Thus the total cover by the first 2,000 GSL words is 70.58 % of the tokens. In terms of types, the first and second 1000 GSL words together account for 17.41% and 8.31%. Although in terms of tokens the GSL covers 64.71 % of the corpus, this accounts for only 17.41% of different word types, indicating a small number of GSL words occur repeatedly in this corpus.

The AWL accounts for 9.13 % of the words in this corpus resulting in the AWL together with the GSL providing a coverage of 79.71 %, which accounts for a little less than four fifths of the total. The remaining 20.29 % of the corpus is made up of non- GSL/ AWL words. These include academic vocabulary specific to the domain, technical words, low frequency words and proper nouns.

Table 1. Coverage of corpus by different word lists

Word lists	Tokens	Percentage in COLC %	Types	Percentage in COLC %
1st 1000 GSL	155206	64.71	2045	17.41
2nd 1000 GSL	14075	5.87	976	8.31
AWL	21905	9.13	1262	10.74
Non GSL/AWL	48664	20.29	7463	63.54

Domain Specific Academic Word List in Chemistry (DAWLIC) and Technical word list compiled from the corpus

The academic words occurring in Chemistry which are not included in Coxhead's AWL were identified using the output from the analysis of the corpus using the software RANGE. A total of 559 words met the criteria for inclusion in DAWLIC. This accounts for 4.76 % lexical coverage.
191 words covering 1.62 % of the vocabulary were identified as technical words.

CONCLUSIONS AND RECOMMENDATIONS

In this study it was shown that the GSL and AWL together accounted for only 79.71 % of the vocabulary in a specialized corpus of language in science. This is consistent with the values obtained in other studies involving science corpora (Valipouri & Nassaji, 2013; Coxhead and Hirsch, 2007). Thus the findings of this study re-emphasize the need to take the academic vocabulary specific to the discipline into account in scientific domains if the lexical threshold is to be achieved. In this situation, the DAWLIC can make an important contribution to English courses catering specifically for science undergraduates.

It is recommended that the contexts of the words in DAWLIC should be examined and they should be taught maintaining the context as discipline specific words may take on different meanings when out of context. As some of the highly specialized meanings of the words in DAWLIC could cause difficulties for the language teacher a two pronged approach involving both language and content teacher would be advisable.

While technical words make a contribution to the vocabulary that cannot be ignored, this is generally considered beyond the purview of the language teacher. One approach to dealing with technical language is the use of subject specific glossaries. However, these are usually vast, covering the entire scope of the subject. This results in poor motivation on the part of the student to use glossaries. In contrast, the use of technical word lists compiled from a corpus has the advantage of giving the most frequent and necessary words only, enabling targeted learning. If the technical words are provided with the equivalent term in the first language the language teacher will be released from explaining scientific principles.

REFERENCES

Chen, Q., & Ge, G. (2007). A corpus based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes, 26*, 502-514.

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

Coxhead, A., & Hirsch, D. (200765-78). A pilot science word list for EAP. *Revue Francaise de Linguistique*.

Laufer, B., & Nation, I. (2011). Vocabulary. In S. M. Gass, & A. Mackey, *The Routledge handbook of second language acquisition* (pp. 163-176). London & New York: Routledge.

Laufer, B., & Ravehorst-Kalovski, G. (2010, April). Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension. *Reading in a foreign language, 22*, 15-30.

Nassaji, H., & Valipouri, L. (2013). A corpus based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*, 248-263.

Nation, P., & Hu, M. (2000). Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, 403-423.

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