



A corpus-based lexical analysis of subject-specific university textbooks for English majors



Konul Hajiyeva*

Department of Linguistics, University of Antwerp, Antwerp, Belgium
Department of Innovations in Teaching, Azerbaijan University of Languages, Baku, Azerbaijan

HIGHLIGHTS

- The study examines eleven subject-specific university textbooks for English majors.
- 2000 high-frequency word families are highly representative in the corpus.
- Academic words constitute a small coverage of the words in the corpus.
- Students need a vocabulary size of 9000 word families to get 98% text coverage.
- A university word list should be developed to meet students' vocabulary needs.

ARTICLE INFO

Article history:

Received 10 June 2015
Received in revised form
12 October 2015
Accepted 15 October 2015
Available online 20 October 2015

Keywords:

Corpus
Word frequency
Academic vocabulary
Text coverage
Word lists
University textbooks

ABSTRACT

This study is a corpus-based lexical analysis of subject-specific university textbooks which purports to explore lexical text coverage and frequency distribution of words from the Academic Word List and the British National Corpus frequency-based word families. For this study a 508,802-word corpus was created, the findings of which reflect that the Academic Word List word families constitute only a small coverage (6.5%) of the words in the entire corpus, whereas the first two thousand high-frequency word families give the coverage of 88.92%. In terms of the text coverage, the results reveal that if 98% coverage of a text is needed for unassisted comprehension, then a vocabulary size of 9000 word families is required. The results also substantiate the claims that the Academic Word List is not as general an academic vocabulary as it was initially intended to be and, more importantly, supports the assumption that students need a more restricted core academic vocabulary. It is therefore argued that 127 academic word families which are relatively frequent in the overall university textbook corpus can be used as a part of the university word list for second-year English majors who have to read and comprehend university textbooks.

© 2015 The Author. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Background

The results of a longitudinal small-scale study of the size of vocabulary mastered by first- and second-year majors in English at the Azerbaijan University of Languages (AUL) show that, after one year of instruction, the extent of their receptive vocabulary usage decreased from 2104 to 1966 word families (Hajiyeva, 2015b). As shown by a similar study of undergraduate Indonesian students undertaken by Nurweni and Read (1999), these students knew half of the General Service List (GSL – a list of frequent lexical items

available for general English) (West, 1953) words, but certainly knew more words (Ward, 2009). The same applies in Azerbaijan, where national secondary school textbooks teach words such as *pilgrimage*, *husbandry*, and *sinister* (Hajiyeva, 2015a). None of these words appear in any of the lists such as the GSL (West, 1953), the Academic Word List (AWL) (Coxhead, 2000) or the University Word List (UWL) (Xue and Nation, 1984). However, taking into account the extent of the vocabulary needed to read and produce academic texts (Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006), I assume, therefore, that the university textbooks (subject-specific manuals of instruction at the faculty of Pedagogy at AUL) to which these students are exposed to in the second year of tertiary education are beyond their comprehension and they do not benefit from them. In other words, as Coxhead et al. (2010, 37) put it forward, 'the reciprocal relationship between vocabulary knowledge and textbooks is critical'. This is because

* Correspondence to: Department of Innovations in Teaching, Azerbaijan University of Languages, Str. R. Behbudov 134, Baku, Azerbaijan.

E-mail addresses: konul.hajiyeva@student.uantwerpen.be, konul_12@rambler.ru.

there is a significant gap between their knowledge of the relevant vocabulary and extent to which it is needed to read academic texts. Thus, taking into account that these second-year students have one year only in which to develop and increase their vocabulary knowledge in order to be able to read and comprehend academic texts, certain steps have to be taken to remedy this situation (Hajiyeva, 2015b).

2. Introduction

Research has identified the breadth of vocabulary knowledge as both a crucial indication of and an important contribution to language abilities, since a rich vocabulary increases mastery of the basic language skills of reading, writing, listening and speaking (Hu and Nation, 2000; Qian, 2002). This is particularly true of reading (Schmitt et al., 2011). Both earlier and recent research studies (Alderson, 2005; Laufer, 1992; Nation and Beglar, 2007) have established a close relationship between reading comprehension and vocabulary knowledge. They have also estimated the number of word families – a ‘word family’ (Bauer and Nation, 1993) includes a single word’s inflections, derivatives and several individual word forms (e.g. *stimulate*, *stimulative*, *stimulation*, *stimulator*, *stimulatory* may all be members of the same word family for a learner with a command of the inflectional suffixes of English) – needed for basic and better comprehension (Laufer, 1992; Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006). For instance, a vocabulary size of approximately 4000 to 5000 word families is that which recent estimates consider as a minimal vocabulary threshold. This represents the minimal vocabulary necessary for reading texts and following lectures at a higher education level (Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt, 2008, 2010; Schmitt and Schmitt, 2012). These recent estimates have also suggested an optimal threshold of 8000 to 9000 word families. This size of vocabulary entails a text coverage of approximately 98% (Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006). Laufer and Ravenhorst-Kalovski (2010, 16) refer to ‘coverage as the percentage of words that a reader understands, for example, if readers have reached 98% text coverage, this means that they understand 98% of the running tokens [words] of the text’.

Unfortunately, most learners from second- or foreign-language contexts do not know or learn this much vocabulary (Schmitt et al., 2011). For example, in her study Laufer (2000) reviewed the vocabulary size of English language learners from eight different countries and reported that this ranged from 1000 to 4000 words or word families, depending on the counting unit. Ward (2009) states that undergraduate students in many countries such as Hong Kong, Indonesia, Jordan, Oman, Sudan and Thailand fall far short of the lexical knowledge necessary to read academic texts in English. These are only a few of the many sources identifying this problem and the same holds true for English majors in Azerbaijan (Hajiyeva, 2014). These first-year English majors do not frequently encounter the great majority of the word families outside the 2000 word band in their general English textbooks and teaching materials since general English textbooks do not provide a good source of words from the most frequent 5000 word families (Hajiyeva, 2015b). The question then is: What is the capacity of such students to comprehend the textbooks and/or teaching materials to which they are exposed? To put it differently, how many words do Azerbaijani English majors need to know to read subject-specific university textbooks?

Determining the vocabulary size of higher education students has had pedagogical consequences, especially informing teachers/lecturers about the number of words to teach, depending on the learning stage of the students (Nizonkiza and Van den Berg, 2014). Acquiring a foreign lexicon is a daunting task for language learners,

especially if the goal is to achieve literacy in the foreign language (Cobb and Horst, 2004). This is also true of those who are learning a foreign language for academic purposes (Chen and Ge, 2007). Specifically for academic purposes, remarkable efforts have been made to assess students’ needs and therefore inform the pedagogy to which they are subjected. Coxhead (2000, 213), for instance, observed that ‘making principled decisions about which words are worth focusing on during valuable class and independent study time’ is the most challenging aspect of vocabulary learning and teaching in an English for Academic Purposes (EAP) programme. Deciding on which words are worth teaching – and which are not – has not been a simple matter (Vongpumivitch et al., 2009) though.

A variety of corpora based on different disciplines were established and analysed in order to improve the learning efficiency of learners of academic vocabulary in an academic environment. As a result, different kinds of word lists were created which considered word frequency as the main criterion. According to Aitchison (1987), word frequency is known to be a factor affecting word familiarity and it is also considered to be a crucial variable in text comprehension. The American University List (Praninskas, 1972) and the University Word List (UWL) (Xue and Nation, 1984) were among the first word lists compiled for academic purposes. Recently, word lists compiled for academic purposes have included Coxhead’s (2000) AWL, Gardner and Davies’s (2013) Academic Vocabulary List and Simpson-Vlach and Ellis’s (2010) Academic Formula List.

A point of controversy has to be noted here: the AWL, for example (Coxhead, 2000), has recently been the subject of some discussion in the literature as to whether a general academic vocabulary really exists. Hyland and Tse (2007) argue that, although Coxhead’s (2000) AWL covers 10% of the academic corpus, it might not be as general academic vocabulary as it was initially intended to be. More specifically, individual items on the list occur and behave in different ways across the disciplines in terms of range (distribution figure), frequency and meaning. The researchers also indicate that teachers and materials developers should not mislead students into believing that there is a single collection of words which they can learn and transfer across different fields of study (Hyland and Tse, 2007; Ward, 2009). It has also been emphasised that students should build a repertoire of specialised academic words in addition to their existing basic or general service vocabulary. Teachers should therefore regard helping their students to establish control of such a vocabulary and to develop a list of high-frequency words, academic and specialised/technical vocabulary in order to guide them through their academic studies. Since readily available lists suffer from the fundamental drawback of not encompassing all the frequent words that certain students are likely to encounter in their reading materials, developing a tailor-made word list may provide direct access to the most frequently used vocabulary for certain disciplines.

Following this discussion, a new AVL (Gardner and Davies, 2013, 1) have been developed from a 120-million-word academic corpus. The authors therefore believe ‘the AVL to be the most current, accurate, and comprehensive list of core academic vocabulary’ since it has been compiled from a much bigger corpus than that of Coxhead’s and it counted lemmas (a headword and some of its inflected and reduced forms) not word families as a counting unit (Gardner and Davies, 2013). Since this core academic vocabulary list is new in the field of academia, empirical studies based on the evaluation of this list are not yet available.

With the above insights in mind, as a starting point, the present study aims to explore lexical text coverage and frequency distribution of words from the AWL (Coxhead, 2000) and Nation’s (2006) frequency lists based on the British National Corpus (BNC) of high-frequency word families in the university textbooks taught at AUL. The BNC is a balanced corpus that represents modern

Download English Version:

<https://daneshyari.com/en/article/1100487>

Download Persian Version:

<https://daneshyari.com/article/1100487>

[Daneshyari.com](https://daneshyari.com)