

ORIGINAL ARTICLE

The Academic Vocabulary List (AVL) coverage in the Brazilian academic written English corpus

Larissa Goulart¹

¹ Northern Arizona University, Arizona, USA.

ABSTRACT

This paper is divided in two parts, in the first one we address the different definitions of academic vocabulary, the role of academic vocabulary in English for Academic Purposes (EAP) teaching, and present some of the academic word lists compiled up to the present day. The second part of this paper is dedicated to one of the applications of these word lists: to determine the vocabulary profile of a corpus. Hence, the investigation conducted relies on the Academic Vocabulary List (AVL) to determine the use of academic vocabulary in a corpus of Brazilian students. It also addresses the issue of the different coverage provided by the AVL and the Academic Word List (AWL). The results indicate that the AVL is more representative of academic vocabulary in the corpus used as a reference.

KEYWORDS: English for academic purposes; academic vocabulary; corpus linguistics.

A cobertura da Lista de Vocabulário Acadêmico no corpus de inglês acadêmico escrito por brasileiros

RESUMO

Este artigo está dividido em duas partes. Na primeira, abordamos as diferentes definições de vocabulário acadêmico, o papel do vocabulário acadêmico em ensino de inglês para fins acadêmicos (IFA), e apresentamos as listas de vocabulário acadêmico compiladas até os dias de hoje. A segunda parte desse trabalho é dedicada a uma das aplicações dessas listas de vocabulário: determinar o perfil lexical de um corpus. Portanto, a investigação conduzida aqui utiliza a Lista de Vocabulário Acadêmico (AVL) para determinar o uso de vocabulário acadêmico em um corpus de alunos brasileiros. Além disso, este trabalho aborda a questão da diferença em cobertura dadas pela AVL e a Lista de Palavras Acadêmicas (AWL). Os resultados indicam que a AVL é mais representativa do vocabulário acadêmico no corpus utilizado como referência.

PALAVRAS-CHAVE: inglês para fins acadêmicos; vocabulário acadêmico; linguística de corpus.

Corresponding Author:

LARISSA GOULART
<lg845@nau.edu>



This article is licensed under a Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original publication is properly cited.
<http://creativecommons.org/licenses/by/4.0/>

1. INTRODUCTION

Academic words are lexical items that are recurrent in academic texts (Chung and Nation, 2003; Clark and Ishida, 2005) and, therefore, are not usually acquired in everyday interaction. Considering English for Academic Purposes (hereafter, EAP) teaching, vocabulary has been surrounded by debates on whether EAP teachers, English teachers who are responsible for teaching academic genres or content area teachers, teachers of a specific field of study, such as Engineering, Law, etc, should focus their classes on it. Several researchers (Flowerdew, 1993; Corson, 1997; Snow, Lawrence, and White, 2009) argue that content teachers might target discipline specific vocabulary for students, whereas they are less likely to do the same with academic vocabulary. In teaching practice, it seems that it is the role of the EAP teacher to teach academic vocabulary. Hence, several researchers (Coxhead, 2000; Gardner and Davies 2014; Browne, Culligan and Phillips 2014, among others) have dedicated themselves to the development of academic word lists that could be used as a resource for EAP teachers and material developers.

Taking this into account, the first part of this paper discusses definitions of academic vocabulary, the role of academic vocabulary in EAP teaching, and presents the academic word lists compiled up to today. The second part addresses one of the applications of these lists: determining vocabulary profile. The academic vocabulary coverage¹ on the corpus of Brazilian students is analysed using the Academic Vocabulary List (AVL) as a reference.

Before discussing previous studies conducted in the area of academic vocabulary, it is important to present some definitions used in these investigations. Firstly, some studies refer to word families, which “consist of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately” (Bauer and Nation, 1993, p. 253). The words *analysed*, *analysers*, *analyses*, *analysing*, *analysis*, for example, are all members of the word family *analyse*. According to Nation (2013), research on vocabulary relies on word families because if students know one or two words in a specific word family they are more likely to understand the other words in that same family. Other studies presented here refer to lemmas, which are “words with a common stem, related by inflection only, and coming from the same part of speech” (Gardner and Davies, 2014).

In addition, most of these studies use the General Service List (GSL) (West, 1953) as a reference, in spite of the criticisms, which say that it is dated (Richards, 1974; Carter, 2012), it lacks in range (Engels, 1968), and it is built based on subjective criteria (Brezina and Gablasova, 2015). The GSL is an inventory of 2,000 high frequency words considered useful for learners of English as a foreign language. Sometimes, it is referred to as the first 2,000 most frequent words in English.

2. WHAT IS ACADEMIC VOCABULARY?

According to Baumann and Grave’s (2010) there is a plethora of explanations associated to academic vocabulary. Li and Pemberton (1994, p.184) state that

¹ Coverage is the amount, in percentage, of a list words present in a text or a corpus. It is what defines the lexical profile of a text.

academic vocabulary is the label given “to this group of words that occur across a range of academic disciplines”. Townsend and Kiernan (2015, p. 113) add to this definition saying that “academic vocabulary words are typically abstract, technical, nuanced, and/or densely packed with meaning”. Some researchers (Hyland and Tse, 2007; Gardner and Davies, 2014), however, claim that a small amount of the words in the Academic Word List (AWL), for example, overlap with the 2,000 most frequent words in English putting into question, then, the distinction between academic and high frequency words. Snow (2010, p. 450), however, argues that “there is no exact boundary when defining academic language; it falls toward one end of a continuum (...), with informal, casual, conversational language at the other extreme.”

For the purposes of this study, it is worth distinguishing general academic vocabulary and discipline specific (academic) vocabulary. General academic vocabulary, also referred simply as academic vocabulary, is a group of words that are frequent in academic texts across different disciplines. On the other hand, discipline specific vocabulary or technical words/terms (Fisher and Frey, 2008; Harmon, Wood and Medina, 2009) are recurrent words in one academic domain, for example, *abdominal*, *laboratory* and *melanoma* are more recurrent specifically in the field of health sciences (Lei and Liu, 2016). Beck, McKeown and Kucan (2002) call discipline specific words Tier 3 words and general academic vocabulary Tier 2 words due to their frequency on academic texts.

McCarthy, O’Keefe and Walsh (2009, p. 7) present the following explanation on English vocabulary:

English often has two kinds of words for the same thing: words whose origin lies in northern Europe (the Nordic and Anglo-Saxon world) and words which came from further south (the Mediterranean world – French, Latin and Greek words). Often, the Greek or Latin word for something is considered more formal than the Anglo-Saxon word for the same thing. Example include *commence* versus *start*, *ascend* versus *go up*, and *depart* versus *leave*.

Words which originate from Greek or Latin are considered more formal inasmuch that around 80% of the academic vocabulary in English comes from Latin or Greek roots (Coxhead and Nation, 2001; Coxhead, 2000; Farid, 1985). This might explain why students whose first languages derived from Latin, such as Portuguese, sometimes find it difficult to perceive if words like the ones presented in the Table 1 are academic words in English or if they are using a Portuguese word with English intonation.

Table 1. Academic words x High frequency words x Portuguese

Academic Word	GSL	Portuguese
commence	begin	começar
obtain	get	obter
indicate	point out/show	indicar

Nevertheless, speakers of Portuguese as a first language are not the only students to whom academic vocabulary might be an obstacle. According

to Vongpumivitch, Huang and Chang (2009) and Li and Pemberton (1994) learners of English as a foreign language have pinpointed vocabulary as one of the most difficult aspects of reading and writing for academic purposes. The studies discussed below provide further evidence on this claim in EAP reading (Park, 2012; Clark and Ishida, 2005) and writing (Shaw, 1991; Townsend et al, 2012).

3. THE ROLE OF VOCABULARY IN LANGUAGE LEARNING

This section will offer an overview into studies on the role of vocabulary in language learning. Several studies (Gass and Selinker, 2008; Cobb and Horst, 2015; Townsend and Kiernan, 2015) argue that having ample lexical knowledge grants better reading and writing skills for students of English as a Foreign Language (EFL). Research on the influence of vocabulary in school-aged children's reading comprehension (Garcia, 1991; Snow, 2010; Stahl and Fairbanks, 1986) confirms the impact of general academic vocabulary in students' achievement. Vocabulary acquisition, however, should be only one among a range of goals addressed in the foreign language classroom (Nation, 2013). Nevertheless, according to Nation (2013) it is possible to approach the skills of listening, speaking, reading and writing and focus on spelling, pronunciation and grammar through the viewpoint of vocabulary. In the EAP context, Townsend et al. (2012, p. 498) remind us that although some studies refer to "academic vocabulary as an independent construct, academic vocabulary, by definition, is academic because of its role within academic language". These authors continue by arguing that although research on academic vocabulary does not guarantee a full picture of students' ability when using academic English, it can be an entry point to the overall academic English proficiency (Townsend et al, 2012, p. 503). Finally, given both these researchers claims, it is the teachers' role to balance the study of vocabulary alongside other aspects of the English language in their classroom practice.

Regarding the correlation between vocabulary size and reading comprehension measures, several studies (Hirsh & Nation, 1992; Hsueh-Cho & Nation, 2000; Nation, 2013) attempt to determine a vocabulary threshold that would allow learners to develop an appropriate understanding of texts. In order to explore this threshold, Hirsh and Nation (1992) investigated the vocabulary size needed to read unsimplified texts for pleasure. These authors analysed the coverage of the GSL in three novels, and their findings suggest that students require a vocabulary of around 5,000 words families in order to read for pleasure. However, the researchers considered that pleasure reading meant encountering one unknown word per 3.3 lines, or knowing 97 to 98% of the words in a text and did not take into account students' actual understanding of the text. Therefore, although this investigation established that the GSL covered only 85 to 90% of the novels, and that students' would have to learn words beyond the GSL in order to understand these narratives, it did not take into account students' lexical knowledge based on reading comprehension tests.

Hsueh-Cho and Nation (2000) explored the relationship between the density of unknown vocabulary and the degree of text comprehension by testing sixty students' comprehension after reading four versions of the same

text with different lexical densities. The first group of students read the complete text, while the second, third, and fourth group of students read the text with, respectively, 5%, 10%, and 15%, of the words changed into nonsense ones. In other words, the fourth group only knew 85% of the words in the text. The results of the reading comprehension tasks suggested that some students in the second group, those who knew 95% of the words, were not able to gain adequate comprehension of the text (Hsueh-Cho and Nation, 2000, p. 415). Therefore, this study confirms the claim that students need to be familiar with 95% or more of the words in a text in order to understand it.

Although these two studies focus on non-academic vocabulary, they corroborate previous researchers' claims that it is necessary to have an extensive vocabulary in order to gain understanding of texts in a foreign language. Li and Pemberton (1994, p. 183), for example, state that if students do not recognise high frequency words they might not be able to carry out top-down reading strategies as they will be employing "too much of the brain's capacity in trying to process this incoming information".

Considering a vocabulary threshold in academic texts, Laufer (1989, p. 317) investigated the correlation between lexical knowledge and the quality of the reading comprehension in academic texts. This researcher asked 100 first year university students to read an academic text and answer comprehension questions. Students were also asked to highlight any unknown words in the text. Her findings sustain the argument that in order for students to understand a text, 95% of the words or more have to be familiar to them. This finding corroborates Coxhead and Nation's (2001) claim that learners of EAP, after acquiring the first 2,000 high frequency words in English, should focus on learning academic vocabulary to gain more text coverage with their vocabulary. According to them, "knowing the 2,000 high frequency words and the words in the academic word list will give close to 90% coverage of the running words in most academic texts. When this is supplemented by proper nouns and technical vocabulary, learners will approach the critical 95% coverage threshold needed for reading" (Coxhead and Nation, 2001, p. 260).

Clark and Ishida (2005) investigated the differences between placed and promoted students² in an advanced EAP in-session course. These authors discovered that, even though promoted students had taken EAP classes and attended lessons in their university courses for one semester, they did not improve in academic vocabulary tests. Furthermore, this finding corroborated teachers' perception that placed students had stronger linguistic foundations than promoted students. In addition, these authors state that "without sufficient knowledge of academic vocabulary, they (students) cannot deal with reading materials efficiently for various types of academic tasks given to them" (Clark and Ishida, 2005, p. 228). This study confirms that academic vocabulary can be an indicator of students' overall academic English proficiency (Townsend et al, 2012).

Park (2012) conducted questionnaires and interviews with foreign students at the same university in which Clark and Ishida's (2005) study

² Placed students are those students who were put in a determined level of English based on a placement test, while promoted students were put in a determined level of English by completing the courses in the previous levels.

was carried out. These interviews showed that students attributed their low reading comprehension ability to their limited vocabulary size (Park, 2012). Furthermore, instructors interviewed for this research also pointed out that students displayed a “difficulty in understanding general academic vocabulary” (Park, 2012, p. 10). Therefore, the instructors interviewed sustain Flowerdew’s (1993) claim that content teachers are less likely to introduce general academic vocabulary to students.

Given the results of the studies presented above, it is clear that EAP students need to have a good command of not only the 2,000 most frequent words in English, but also of academic and technical vocabulary. Nevertheless, understanding vocabulary is not the same as producing appropriate vocabulary for academic texts, as students might be able to understand academic words but might not be able to use them in their own texts. The studies below present how students deal with academic vocabulary production.

Shaw (1991) interviewed 22 foreign students who were writing their theses or dissertations in English. These students declared that “vocabulary and finding the right word for the context” (p. 195) were their most significant problem when writing EAP. The same results were obtained by Leki and Carson (1994) who interviewed 128 foreign students.

Regarding the impact of vocabulary knowledge in academic achievements, Townsend et al (2012, p. 502) investigated the correlation between academic success and general academic vocabulary knowledge in a group of 339 students. These students took a general academic vocabulary test and the results were compared with their academic achievements. Students who scored higher in the vocabulary test also achieved higher marks in their content tests. Furthermore, Townsend et al. (2012) analysis took into account students’ linguistic background (English only or EFL). English-only students outperformed students whose first language was not English, supporting Li and Pemberton’s (1994) claim that academic vocabulary is more problematic for non-native students.

Engber (1995), in his study of ESL students writing in an academic context in the USA, found that students who incurred in high amounts of lexical errors presented lower lexical density and lower lexical variation received worse grades. Thus, corroborating Santos’s (1988) earlier findings that professors of the content area judge lexical errors as the most serious of students’ language errors, and they are more likely to penalize students’ final grade based on lexical errors than other language mistakes.

The objective of this section was to present previous studies that investigated the vocabulary threshold for general and academic purposes and studies on students’ difficulties with reading and writing for academic purposes. The next section presents arguments in favour and against teaching general academic vocabulary.

4. WHY TEACH ACADEMIC VOCABULARY?

Coxhead and Nation (2001) claim that general academic vocabulary should be emphasized in course plans and teaching materials because: a) academic vocabulary is common to a wide range of academic texts and generally not

so common in other texts; b) academic vocabulary accounts for a significant number of words in academic texts; c) academic vocabulary is generally not as well-known as technical vocabulary, and d) English teachers can usefully help the learners with academic vocabulary, while this might not be true for technical vocabulary, in which the teacher probably does not have the necessary background to understand the words. Hyland and Tse (2007) and Durrant (2014), however, question the usefulness of a general academic word list. Hyland and Tse (2007) criticize especially argument “b” which says that academic vocabulary accounts for a large portion of words in an academic text. According to these researchers, it would be more helpful if students learned discipline specific vocabulary rather than general academic vocabulary. They claim that a discipline specific vocabulary would cover larger portions of the texts and, therefore, would promote better reading comprehension and, as a consequence, active use when writing academic texts. Nevertheless, Trimble (1985) claims exactly the opposite. According to Trimble (1985), discipline specific vocabulary is not enough of a problem for students of EFL, due to its low frequency in academic texts. Corroborating Trimble’s (1985) argument, previous investigations (Campion and Elly, 1971; Praninskas, 1972) uphold the existence of a compendium of academic words that perform a supportive role in academic texts. Besides, previous studies conducted on the coverage of academic words in discipline specific corpora, such as Chung and Nation’s (2003) research on anatomy and applied linguistics textbooks and Chen and Ge’s (2007) study on a corpus of medical research articles confirm the supportive role of academic vocabulary in academic texts. Furthermore, considering the applications of vocabulary compendiums into teaching, most EAP classrooms are not discipline specific (Durrant, 2016), therefore a general academic word list might be more useful as a resource for English teachers. Finally, as Coxhead (2011, p. 357) points out, “a number of subject-specific vocabulary lists have been developed recently to address the needs of particular learners (e.g., Wang, Liang, & Ge, 2008; Coxhead & Hirsh, 2007; Ward, 2009; Chung, 2009)” among other lists compiled after the publication of Coxhead’s article, such as Lei and Liu’s (2016) medical academic word list and Yang’s (2015) nursing academic word list. These lists can be used to supplement the vocabulary of EAP students with regard to their specific area of study/research.

Hyland and Tse (2007) also question the suitability of a general academic word list to all fields of expertise based on the fact that even if students were to acquire a general academic vocabulary, words might assume different meanings according to their fields of study. They exemplify this claim with the word “attribute”, which is recurrent in linguistics as the verb “to accredit” and in economics as the noun “feature” (Hyland and Tse, 2007, p. 245). Nevertheless, according to Nagy and Townsend (2012, p. 96) this feature is part of the nature of academic vocabulary, their “dictionary entries tend to include many definitions”. In addition, Townsend et al. (2012) point out that meaning differentiation also occurs in high frequency words. Therefore, teachers can make these distinctions salient to students either by using specific vocabulary lists in a supportive role or even with students searching for occurrences of the academic words in texts from their areas of study.

Finally, as mentioned previously some authors suggest that content teachers can be responsible for teaching discipline specific vocabulary, while the EAP teacher is responsible for teaching general academic vocabulary. Other researchers (Cowan, 1974; Hutchinson and Waters, 1987; Farrell, 1990) also sustain that EAP teachers might not be able to target this discipline specific vocabulary and that EAP learners would not encounter difficulties with this vocabulary. These claims support the view presented on argument “d” which mentions that English teachers can help learners with a general academic vocabulary and might not be as helpful with technical vocabulary.

Taking into consideration the need for EAP teachers to teach academic vocabulary, many researchers have taken upon themselves to create lists of academic vocabulary that could be used as a resource for teachers, material developers, test item developers, and students. The following section presents these word lists.

5. ACADEMIC WORD LISTS: A TIMELINE

The objective of this section is to discuss the evolution of academic word lists throughout the years, taking into consideration the methodology used to create these lists, and their applications. Additionally, this section presents the differences between the AVL and the AWL, which are the two main academic word lists present in research nowadays. The table, below, presents nine general academic word lists compiled from 1971 (Campion and Elley) to 2014 (Gardner and Davies; and Browne, Culligan and Phillips).

The first point to consider regarding this table is that almost all of the word lists presented here, with the exception of Lynn (1973), Ghadessy (1979) Paquot (2010), and Gardner and Davies (2014), exclude high frequency words whether by eliminating GSL words or words in other general lists, such as the New-GSL or the most frequent words in the *Teacher's book of words* by Thorndike and Lorge (1944). However, Paquot (2007) questions the exclusion of GSL words, findings in her investigation indicate that two thirds of the nouns in her Academic Keyword List (AKL) are also in the GSL, while only one third of these nouns occur in the AWL. In addition, Gardner and Davies (2014) argue that it is not useful to distinguish between academic and high-frequency words, as academic words make an additional contribution to general vocabulary. Therefore, these authors chose not to associate the AVL with any general English word list.

Furthermore, Lynn's (1973) and Ghadessy's (1979) take into account students' annotations and translations written in their textbooks to compile their word list, therefore these lists do not omit GSL words as these words might be relevant to those students. The methodology adopted to build these two lists is somewhat questionable since students' annotations do not necessarily mean that they did not understand the words. Another aspect to consider is that Lynn's (1973) word list was intended specifically for the context of students at Nanyang University in Singapore. Nevertheless, it was used as a reference, together with Champion and Elley's (1971) Academic Vocabulary List, Praninskas' (1972) American university word list and Ghadessy's (1979) word list, to form the University Word List (UWL, 1984). Therefore, although the UWL was widely used in vocabulary research

before Coxhead's AWL, it also combined the methodological weaknesses of all the previous lists. Coxhead's (2000) AWL was the first general academic vocabulary compendium to rely on computer tools in order to determine frequency, range and distribution on its compilation.

A second point to be noted is the use of non-professional writing in the reference corpus. Paquot (2010) and Browne, Culligan and Phillips (2014) were the only authors to include students' writing in the corpus that originated their word lists. Nesi (2008, p. 03) argues that "novice writers do not begin by writing for publication, and their early attempts at academic writing are likely to be assessed texts produced in the context of a course of study". Therefore, pointing out to the need to consider the texts of novice writers in the production of academic word lists. Browne, Culligan and Phillips (2014) were also the first researchers to include spoken academic English in their reference corpus.

Finally, we have to consider the dichotomy between word families and lemmas in recent vocabulary lists. Although word families have been criticised for not taking into account parts of speech and, also, because derived and inflected words are not always related in meaning to the headword in a family (Gardner and Davies, 2014), they are extensively used in the production of word lists, as we can perceive in Table 2. Even the AVL, originally composed of lemmas, has a version in word families. Browne, Culligan and Phillips (2014) are the only authors not to make available a word family or word form version of their academic word list. Nevertheless, they state on their website that this list is still under development, hence there is a possibility that in the future a version with word families, or word forms will be made available. It is also possible to consider that recent word lists are also available as word families due to the easiness to explore word families in a corpus.

Table 2. Academic word lists timeline

Author/Year	Name	Outcome	Procedure	Applications	Access
Campion and Elley, 1971	Academic Vocabulary List	Two lists, one containing the complete academic vocabulary with 3,200 words and the other containing the top 500 words in the academic vocabulary.	This list was compiled based on a frequency word count of university textbooks and high school students rating of these words difficulty if encountered in a text. The final list also excludes the 5,000 most frequent words in English according to Thorndike and Lorge (1944).	This list used to be the reference for the academic section of the Vocabulary Level Test (Nation, 1983).	
Praninskas, 1972	American University Word List	This list contains 507 word families organized in alphabetical order, frequency of each word form or frequency of word families.	It was compiled based on the texts of ten textbooks used in first year undergraduate classes in the courses of chemistry, biology, psychology, physics, English, literature, sociology and history at the American University of Beirut. The manual corpus used to create this list contained 30,844 running words. The final vocabulary lists excludes proper nouns, dates, foreign words, abbreviations, discipline specific words and words that appeared in the GSL.	Textbooks (Yorkey, 1981; Farid, 1985; Valcourt and Wells, 1999).	

continua

Table 2 (continuação)

Author/Year	Name	Outcome	Procedure	Applications	Access
Lynn, 1973	Word list	This list contains 10,000 types organized according to frequency of occurrence.	In order to compile this list, Lynn (1973) checked 52 textbooks and 4 cycle-styled handouts used at the University of Nanyang looking for the words which students annotated or wrote the translation in their academic texts.		
Ghadessy, 1979			Ghadessy (1979) also looked into student's annotations on textbooks, examining a manual corpus of student's textbooks with more than 478 running words.		
Xue and Nation, 1984	A University Word List (UWL)	This list contains 836 word families	Xue and Nation (1984) combined the items on the four previous word lists (Campion and Elly, 1971; Praninskas, 1972; Lynn, 1973 and Ghadessy 1979), excluding the words in the GSL.	This list used to be the reference for <i>Range</i> and the <i>VocabProfile</i> tool in LexTutor.	Available at https://www.learnthat.org/word_lists/view/7797
Coxhead, 2000	A new Academic Word List (AWL)	A list containing 570 word families divided into 10 frequency-ranked sublists. Therefore, the first sublist accounts for one third of the words in an academic corpus, while the second sublist accounts for half of this amount. Each sublist, with the exception of the last one, contains 60 words families.	This list was compiled based on an academic corpus of 3.5 million words. The texts in this corpus included the areas of arts, commerce, law, and science. The criteria to include the words in the AWL were: frequency – the words had to occur more than 100 times; range - the words had to occur at least 10 times in each subcorpora; and specialised occurrence – the words could not be part of the GSL.	– The Vocabulary Level Test (Schmitt, Schmitt and Clapham, 2001). – LexTutor - Vocabprofile (Cobb, n.d.). – Textbooks (Savage and Mackey, 2010; Huntley 2006; Zimmerman et al, 2012; Schmitt and Schmitt, 2011; Mifflin, 2006). – Oxford Student's Dictionary. – Longman Exams Dictionary.	Available at http://www.victoria.ac.nz/lals/resources/academicwordlist/
Paquot, 2010	Academic Keyword List (AKL)	The AKL contains 930 potential academic words divided into nouns, verbs, adjectives, adverbs, and others.	This list was based on two corpora of professional writing – Micro-Concord corpus collection B (Scott and Johns 1993) and the Baby BNC Academic Corpus – and two corpus of student writing – the Louvain Corpus of Native Speaker Essays and the British Academic Written English Pilot Corpus. The criteria to select the words in this list were: keyness – the words had to appear significantly more in the academic corpus than in a comparison corpus of fiction; frequency, range, and distribution.	Louvain EAP Dictionary (Granger and Paquot, 2010).	Available at https://www.uclouvain.be/en-372126.html
Gardner and Davies, 2014	A new Academic Vocabulary List (AVL)	This is a list of the 3,000 top lemmas occurring in all academic domains of the Corpus of Contemporary American English. A version of this list organized in word families for research and teaching purposes is also available.	Gardner and Davies' (2014) AVL derives from 120 million words subcorpus of academic English in Corpus of Contemporary American English (COCA). The words were included in this list based on: ratio – words had to occur at least 50% more in the academic subcorpus than in other non-academic subcorpora; range – words had to occur in seven of the nine academic disciplines with more than the expected frequency; dispersion – the word had to occur with a dispersion above 0.80; discipline measure – the word could not occur more than three times the expected frequency in any of the disciplines.		Available at http://www.academicvocabulary.info/

continua

Table 2 (conclusão)

Author/Year	Name	Outcome	Procedure	Applications	Access
Browne, Culligan and Phillips, 2014	New Academic Word List (NAWL)	The NAWL contains 963 lemmas derived from a 288 million-word corpus. This list is available in three different forms: lemmas, headwords only and based in frequency indices.	The corpus used to compile this list incorporated the academic subsection of the Cambridge English corpus, best-selling textbooks, the MICASE corpus and the BASE corpus. Words were selected based on frequency, dispersion and appropriateness. In addition, words in the New General Service List (NGSL, Browne, Culligan and Phillips, 2014) were excluded.	– LexTutor – Complete Vocab Profile (Cobb, n.d.). – New Academic Word List Test (Bennett and Stoeckel, 2015).	Available at http://www.newacademicwordlist.org/

Despite the fact that the AVL has had more applications in the development of textbooks, tests and research on vocabulary, the second part of this study relies on the AVL as a reference for academic vocabulary based on the reasons presented next. Firstly, the AVL has been criticised for being dated (Gardner and Davies, 2014) and especially because it was compiled on top of the GSL (West, 1953). Considering the use of the GSL, which is considered an outdated list of high frequency words, Gardner and Davies (2014) argue that the AVL covers a high percentage of academic texts because it contains a considerable amount of modern high frequency words. In other words, according to these authors, the AVL would cover academic and non-academic texts alike.

Another major criticism towards the AVL is the fact that it was created based on word families (Nagy and Townsend, 2012). Gardner and Davies (2014) acknowledge the fact that word families are easier to be used as a reference when conducting research, therefore the AVL contains a word family form, nevertheless, these authors argue that the issue with the AVL is that it was compiled based on the word families, rather than being compiled based on lemmas and then converted into word families. Furthermore, the AVL also contains word frequency and parts of speech information in its word family form, which does not occur in the AVL. In addition, the reference corpus used to create the AVL contained more recent texts, thus, being representative of current academic language and it contained more than 120 million words, while the corpus of reference for the AVL contained 3.5 million words. Considering this, it is worth assessing the vocabulary profile in the Brazilian Academic Written English (BrAWE, Silva, 2017) corpus using the AVL as a reference.

Hence, the second part of this paper uses corpus linguistics tools to investigate the coverage of academic vocabulary in BrAWE. This corpus comprises texts of assignments written by Brazilian students during their exchange programmes at British universities. It was created to be comparable with the British Academic Written English Corpus (BAWE). These texts were written as part of students' assignments in the undergraduate course they were taking. This corpus is divided in the same subcorpora as BAWE, that is: four areas of expertise – Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences, and 13 genre families – Case Study, Critique, Design, Empathy Writing, Essay, Exercise, Explanation, Literature Survey, Methodology Recount, Narrative Recount, Problem Question, Proposal, and

Research Report. BrAWE has 670,314 tokens distributed in 380 text files written by 225 students.

Finally, some researchers (Gardner and Davies, 2014; Olsson, 2015; and Newman, 2016) have already compared the coverage of AVL and AWL in different academic corpora, in order to verify if the AVL is an appropriate substitute to the AWL. The same procedure will be done here, comparing the results of the AVL coverage in BrAWE with the AWL coverage in BrAWE (Silva, Matte and Sarmento, 2018). Taking this into account, the proceeding sections seek to answer the following questions:

- a) How is the AVL coverage in BrAWE compared to the AVL coverage in other academic corpora previously investigated, namely, Gardner and Davies (2014), Olsson (2015), Durrant (2016), and Newman (2016)?
- b) How does the coverage of the AVL differ from the coverage of the Academic Word List (AWL) in the BrAWE corpus?

In the next section other studies that have determined the coverage of the AVL in academic corpora are presented.

5. ACADEMIC VOCABULARY COVERAGE IN ACADEMIC CORPORA

This section discusses the results of previous studies which have explored the AVL coverage in written academic corpora and the AWL coverage in BrAWE.

The first researchers to explore the coverage of the AVL were its compilers. Gardner and Davies (2014) studied the coverage of the top 570 word families in the AVL in the academic subcorpora of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These authors chose to analyse only the 570 top word families and not the lemmatized version of their list in order for their results to be comparable with the AWL coverage of the same corpora. In the Academic subcorpus of the BNC, the AVL accounts for 13.7% of all the tokens, and in COCA the AVL covers 13.8%. To validate the AVL, these authors have also compared its coverage in non-academic corpora, which resulted in a significant lower figure, as can be seen in Table 3. This result shows that the AVL is relevant, specifically, for the study of academic English, and not only a list of high frequency words that occur in any kind of text.

Table 3. AVL coverage in COCA and BNC (Gardner and Davies, 2014, p. 323)

Genres	COCA			BNC		
	Genre size	# Words AVL	Coverage	Genre size	# Words AVL	Coverage
Academic	120,847,709	16,633,796	13.8%	32,828,961	4,507,211	13.7%
Newspaper	77,553,000	6,229,359	8.0%	10,638,034	740,065	7.0%
Fiction	83,369,907	2,862,093	3.4%	16,194,885	548,708	3.4%

These authors also compared the AWL coverage in both academic corpora, coming to the conclusion that the AWL coverage is almost half – 7.2% in COCA and 6.9% in BNC – of the AVL coverage in the same corpora, as can be seen in the table four, below.

Table 4. AVL and AVL coverage in COCA and BNC (Gardner and Davies, 2014, p. 323)

List	COCA Academic			BNC Academic		
	Genre size	# Words	Coverage	Genre size	# Words	Coverage
AVL (570)	120,847,709	16,633,796	13.8%	32,828,961	4,507,211	13.7%
AWL (570)	120,847,709	8,601,839	7.2%	32,828,961	2,261,469	6.9%

Olsson's (2015) study aimed at comparing the development of academic vocabulary in students in a CLIL (Content and Language Integrated Learning) classroom. Therefore, he has analysed the AVL coverage in texts written during four stages of students' learning. In their final stage – considered the most proficient students, the AVL represented 12.14% of the words in the corpus. Olsson (2015) has conducted the same research using the AWL as a reference for academic vocabulary. In this corpus, the AWL covers 4.94% of words, less than half of the words accounted for in the AVL. As a conclusion, this author states that the AVL shows the development of students clearly, while the AWL results do not indicate students improvement.

Newman's (2016) study compared the coverage of the AWL and the AVL in the Academic Textbook Corpus (ACT). In the same way as Gardner and Davies (2014), Newman (2016) compared the 570 word families represented in the AWL to the top 570 word families in the AVL. The results indicate an overall coverage of 13.4% by the AVL in the ATC corpus, while the AWL accounted for 6.67% of the words in the corpus. The results also highlight other differences between the AVL and the AWL. According to Newman (2016, p. 26) "AVL types appeared more consistently across grade levels in the corpus" and also the 570 word families in the AVL were found in the corpus, while 567 word families in the AWL were found in the corpus.

The last study that has analysed the vocabulary profile of an academic corpus using the AVL as a reference is Durrant (2016). This researcher has evaluated the AVL coverage in BAWE. Durrant's (2016) results show that the AVL covers 16.82% of all words in the corpus. As a result of this study, Durrant (2016) suggests an Academic List comprising the 427 AVL items which were frequent across 90% of the academic disciplines represented in BAWE. This list is, thus, useful for EAP teachers and students as it is a shorter version of the complete AVL representing only high frequency words in student academic writing. The Table 5 summarizes the findings of the four investigations presented above.

Table 5. AVL and AWL coverage in previous research

Study	AVL Coverage	AWL Coverage
Gardner and Davies (2014) - BNC, Academic.	13.7%	6.9%
Gardner and Davies (2014) - COCA, Academic	13.8%	7.2%
Olsson (2015)	12.14%	4.94%
Newman (2016)	13.4%	6.67%
Durrant (2016)	16.82%	
BAWE	14.3%	9.5%

One of the main differences between the study conducted by Durrant (2016) and other studies concerning the AVL coverage is that this author has used the lemmatized version of the list, which reflected in an increase coverage of the AVL, as can be seen in Table 5. Taking this methodological difference into account, the author of this paper has analysed the coverage of the 570 top word families of the AVL in BAWE, as well as the AWL list. The results of this analysis are presented in the last line of the table.

As we can see from the table above, in studies which have used the word families version of the AVL as a reference, the academic vocabulary coverage varies between 12.14 to 14.3%. Therefore, considering that BrAWE is a learner corpus, the AVL coverage in this corpus is expected to fall either within these numbers or below them. In addition, we can notice that the AWL coverage, when available, was significantly lower.

The aim of this section was to present preceding investigations on the AVL coverage in academic corpora. The next section briefly addresses a previous study on vocabulary profile conducted in BrAWE which used the AWL as a reference.

6. PREVIOUS STUDY ON BRAWE

Silva, Matte and Sarmento (2018) have analysed the vocabulary profile of Brazilian students using the AWL and the GSL as a reference. These authors opted for these lists because many investigations have been conducted on the AWL coverage in different academic corpora and discipline specific corpora, which has not taken place with the AVL yet. Therefore, using the AWL they were able to compare the academic words coverage taking into account areas of expertise.

Table 6. AWL coverage in BrAWE (Silva, Matte and Sarmento, 2018, p. 515)

	AH (7.887)			SS (67.907)			LS (224.979)			PS (369.541)			Total (670.314)		
	GLS (%)	AWL (%)	OFF (%)	GLS (%)	AWL (%)	OFF (%)	GLS (%)	AWL (%)	OFF (%)	GLS (%)	AWL (%)	OFF (%)	GLS (%)	AWL (%)	OFF (%)
Case Study				78.60	9.70	11.70	74.10	8.90	17.00	79.00	10.40	10.50	77.60	9.90	12.50
Critique				79.70	11.00	9.40	77.00	11.00	12.00	77.20	11.10	11.80	77.60	11.00	11.40
Design										78.70	10.40	10.80	78.70	10.40	10.80
Empathy Writing															
Essay	81.70	5.20	13.10	80.60	9.40	10.00	69.50	9.60	20.90	77.30	10.50	11.30	73.90	9.60	16.50
Exercise				86.40	5.20	8.40	70.80	8.40	20.80	78.70	9.90	11.40	77.60	9.50	12.90
Explanation							73.00	8.00	19.10	78.60	8.90	12.50	77.90	8.90	13.20
Literature Survey							71.00	9.30	19.80	62.40	7.50	30.10	69.30	8.90	21.80
Methodology Recount							69.30	8.20	22.50	79.20	9.40	11.30	75.60	8.90	15.50
Narrative Recount							74.50	11.80	13.80	80.90	10.40	8.60	78.40	10.90	10.60
Problem Question							78.60	12.40	9.10	78.50	7.80	13.70	80.50	9.70	9.80
Proposal							76.30	11.30	12.40	77.80	12.40	9.80	77.40	12.20	10.30
Research Report							69.10	9.10	21.80	75.40	10.70	13.90	73.10	10.20	16.70
Total	81.70	5.20	13.10	80.30	9.80	9.90	71.30	9.40	19.30	77.90	10.20	12.00	76.00	9.80	14.20

Their results, presented in Table 6, above, show that the AWL amounts for 9.8% of the words in the corpus, with some discipline variation – 5.2% in Arts and Humanities and 10.2% in Physical Sciences.

7. AVL COVERAGE IN BRAWE

Range (Nation and Heatley, 2002) was used to determine the AVL coverage of BrAWE. This software allows for up to 32 text files to be analysed based on different word lists. According to Nation and Heatley (2002) “for each word in the texts, it (Range) provides a range or distribution figure, a headword frequency figure, a family frequency figure, and a frequency figure for each of the texts the word occurs in”. Although Range comes with the GSL and the AVL built in, it is possible to alter or add up to 10 word lists in its system. For this study the AVL was used as the only base word list in the software.

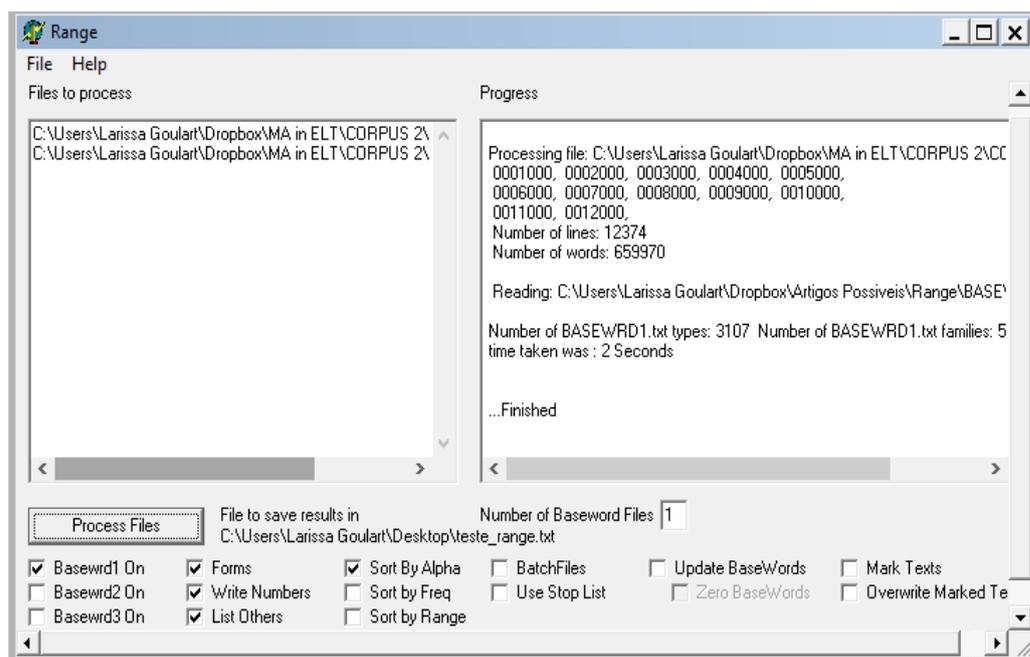


Image 1. Range Screenshot

The AVL in word families is available at <https://www.academicvocabulary.info>. In this website it is possible to download the 3,000 top lemmas in the AVL, the AVL in word families, and the 20,000 most frequent words in COCA. For this study, in the same way as Newman (2016), Gardner and Davies (2014), and Olsson (2015) the top 570 word families were used as a reference. After downloading the list from the academic profile website, it was necessary to convert it, to clean it from other information – such as frequency, area of expertise associated, and part of speech – and convert it to txt. Table 7 presents the results divided on BrAWE subcorpora the same way as Silva, Matte and Sarmento (2018) presented their results.

Table 7. AVL coverage in BrAWE

	AH*	SS	LS	PS	Total
	AVL	AVL	AVL	AVL	AVL
Case Study		13.5%	14.4%	15.2%	14.7%
Critique		17.6%	17.1%	16.6%	17.0%
Design				14.3%	14.3%
Empathy Writing					
Essay	8.9%	13.6%	14.3%	15.7%	14.4%
Exercise		10.6%	13.0%	15.9%	15.2%
Explanation		14.9%	12.8%	13.5%	13.6%
Literature Survey			13.9%	10.6%	13.3%
Methodology Recount			13.6%	15.0%	14.5%
Narrative Recount			22.4%	14.5%	17.5%
Problem Question		10.7%	18.8%	12.0%	14.2%
Proposal			16.2%	18.6%	18.1%
Research Report			14.1%	15.4%	15.0%
Total Tokens	8.9%	14.5%	14.5%	15.2%	14.8%

* AH stands for Arts and Humanities, SS for Social Sciences, LS for Life Sciences and PS for Physical Sciences.

Taking into account word families, the 570 word families in the AVL were represented in BrAWE, while 566 word families from AWL were present in BrAWE. As we can see, the total coverage of the AVL in BrAWE is 14.8%, with some variation in genres of discourse and area of expertise. Arts and Humanities is the area of expertise with the lowest use of academic vocabulary. Nevertheless, as Silva, Matte and Sarmento (2018) point out it is impossible to draw conclusions from the Arts and Humanities subcorpora as it is notably underrepresented in BrAWE. Considering the different genres, it is possible to notice that Proposal (18.1%) and Narrative Recount (17.5%) use substantially more academic words than other genres, such as Literature Survey (13.3%).

The aim of this section was to present the AVL coverage in BrAWE. The next section will discuss the results and answer the research questions, as well as examine the limitations of this study.

8. DISCUSSION

The aim of this section is to discuss the results of this investigation in light of the literature review. Taking into account that vocabulary is considered one of the main issues in EAP learning (Gass and Selinker, 2008; Cobb and Horst, 2015; Townsend and Kiernan, 2015), many authors have set to compile word lists that could be useful for students to evaluate their own texts; and for teachers when considering which words to focus their teaching on. In the second part of this study, one of those lists was used as a reference in a corpus of Brazilian students' writings. This analysis shows one of the applications of wordlists while it also highlights the fact that students represented in the BrAWE corpus are not, at least considering AVL coverage, less apt to use academic vocabulary than users of EAP represented in

other academic corpora. The research questions proposed in the second part of this study address the quantitative aspects of comparison between the AVL and the AWL, they will be answered separately in the next paragraphs.

- *How does the AVL coverage in BrAWE compare to the AVL coverage in other academic corpora previously investigated, namely, Gardner and Davies (2014), Olsson (2015), Durrant (2016), and Newman (2016)?*

Previous studies that have investigated the AVL coverage, such as, Gardner and Davies (2014), Olsson (2015), Newman (2016) and Durrant (2016) have obtained a coverage ranging from 12.4% to 16.8%. Nevertheless, these studies had methodological differences, Durrant (2016), for example, used the lemmatized version of the AVL, hence explaining its increased coverage. When analysing the same corpus with the top 570 word family version of the AVL this author has found it to cover 14.3% of BAWE. Another major difference is that Durrant (2016) and Olsson (2015) have used a student writing corpus as a reference, while Gardner and Davies (2014) and Newman (2016) have used a corpus of professional writing corpus. This, however, did not show significant coverage variation. The issue of different coverages in student versus professional corpus is something that will need to be taken into account as more studies concerning the AVL coverage in academic corpora are conducted.

Students in BrAWE surpass those in previous studies when considering the coverage of academic words in their texts, as they reached a 14.8% of words in the AVL list. Nevertheless, this result should be taken as only the quantitative part of this analysis. It is still necessary to analyse qualitatively the words selected by these students. In previous investigations of academic vocabulary (Silva, 2016) coverage, qualitative analysis has shown that even though students use the same amount of academic vocabulary as other professional academic corpora, they tend to repeat the same words, indicating that they had lower lexical density. Future studies in BrAWE should address this question.

- *How does the coverage of the AVL differ from the coverage of the Academic Word List (AWL) in the BrAWE corpus?*

Almost all previous research concerning the AVL coverage in academic corpora have also compared the coverage of the AWL with the AVL results. Therefore, this study follows the same methodology. In these previous studies the AVL represent more words than the AWL in academic corpora, amounting to almost double the coverage of the AWL in Gardner and Davies (2014), Olsson (2015) and Newman (2016). According to Gardner and Davies (2014) this happens because the AVL comprises more words that are specific from academic texts, whereas the AWL by excluding the words in the GSL, also excluded relevant academic words. It is worth pointing out that the AVL is one of the few word lists which did not exclude high frequency word lists, such as the GSL or the New-GSL, in its compilation.

Although the results of this study reinforce previous findings by showing that AVL covers more (14.8%) of the BrAWE corpus than the AWL (9.6%), it is surprising that the amount increased is not as significant as the ones shown

in previous studies. However, the same can be observed when taking into account the AVL and the AWL coverage of BAWE, which might indicate that student academic writing might present less significant difference.

One issue that might be worth noting is that while the AWL accounts for both British and American spelling of the words, the AVL not always contains both forms for the same word. Therefore this might have influenced the results.

This study is a preliminary investigation of the AVL coverage in a corpus of Brazilian EAP. It showed that Brazilian students analysed in the sample have used academic vocabulary as other professional corpora. Nevertheless, it points out the need for further research into the academic words selected by these students, as well as for investigations on the coverage of academic words in different academic genres.

Finally, considering the role of academic vocabulary and the need to connect research and the classroom, future investigations could analyse the coverage of all most recent academic lists, such as, the AWL, AKL, AVL and the New-AVL in a professional corpus of EAP and investigate the resources available for teachers and students of EAP with each of these lists. The AVL for example, has its own website where not only the list is available, but also other teaching resources.

REFERENCES

- Bauer, L. & Nation, I. S. P. 1993. Word families. *International Journal of Lexicography*, 6, p. 253-279.
- Baumann, J. & Graves, M. 2010. What is academic vocabulary? *Journal of Adolescent and Adult Literacy*, 54(1), p. 4-12.
- Beck, I., McKeown, M., & Kucan, L. 2002. *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Brezina, V. & Gablasova, D. 2015. Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36, p. 1-22.
- Browne, C., Culligan, B. & Phillips, J. 2014. *A New General Service List*. Retrieved from <<http://www.newgeneralservicelist.org/>>.
- Campion, M. & Elley, W. (1971). *An Academic Word List*. Wellington New Zealand Council for Educational Research
- Carter, R. 2012. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- Chen, Q. & Ge, C. 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, p. 502-514.
- Chung, T. & Nation, P. 2003. Technical vocabulary in specialised texts. *Reading in a foreign language*, 15(2), p. 103-116.
- Chung, T. 2009. The newspaper word list: A specialised vocabulary for reading newspapers. *JALT Journal*, 31(2), p. 159-182.
- Clark, M & Ishida, S. (2005). Vocabulary knowledge differences between placed and promoted EAP students. *Journal of English for academic purposes*, 4, p. 225-238.
- Cobb, T. & Horst, M. 2015. Learner Corpora and Lexis. In. S. Granger, G. Gilquin & F. Meunier (Eds.). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

- Corson, D. 1997. The learning and use of academic English words. *Language Learning*, 47(4), p. 671-718.
- Cowan, J. R. 1974. Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), p. 389-400.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly*, 34 (2), p. 213-238.
- Coxhead, A. & Nation, P. 2001. The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.). *Research perspectives on English for academic purposes* (p. 252-267). Cambridge: Cambridge University Press.
- Coxhead, A. & Hirsh, D. 2007. A pilot science word list for EAP. *Revue Francaise de linguistique appliquee*, XII(2), p. 65-78.
- Coxhead, A. 2011. The Academic Word List 10 years on: research and teaching implications. *TESOL Quarterly*, 4(2), p. 355-362.
- Durrant, P. 2014. Discipline and Level Specificity in University Students' Written Vocabulary. *Applied Linguistics*, 35(3), p. 328-356.
- Durrant, P. 2016. To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, p. 49-61.
- Engber, C. A. 1995. The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), p. 139-155.
- Engels, L. 1968. The fallacy of word-counts. *International Review of Applied Linguistics in Language Teaching*, 6(3), 213-231.
- Farrell, P. 1990. Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary. *CLCS Occasional Paper No. 25* Trinity College.
- Farid, A. 1985. *A vocabulary workbook: Prefixes, roots, and suffixes for ESL students*. Englewood Cliffs, NJ: Prentice Hall.
- Fisher, D. & Frey, N. 2008. *Word wise and content rich: five essential steps to teaching academic vocabulary*. Portsmouth, NH: Heinemann.
- Flowerdew, J. 1993. Concordancing as a tool in course design. *System*, 21, p. 231-244.
- Garcia, G. 1991. Factors influencing the English reading text performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, 26(4), p. 371-392.
- Gardner, D. & Davies, M. 2014. A new academic vocabulary list. *Applied Linguistics*, 35(3), p. 305-327.
- Gass, S. & Selinker, L. 2008 *Second language acquisition: an introductory course*. New York: Routledge/Taylor and Francis Group.
- Ghadessy, P. 1979. Frequency counts, words lists, and materials preparation: A new approach. *English Teaching Forum*, 17, p. 24-27.
- Harmon, J. M., Wood, K. D., & Hedrick, W. B. 2009. Vocabulary instruction in middle and secondary content classrooms: Understandings and direction from research. In A. E. Farstrup & S. J. Samuels (Eds.). *What research has to say about vocabulary instruction* (p. 150-181). Newark, DE: International Reading Association.
- Hirsh, D. & Nation, P. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), p. 689-696.
- Hyland, K. & Tse, P. (2007). Is there "an academic vocabulary"? *TESOL Quarterly*, 41(2), p. 235-253.
- Hsue-Cho, M. & Nation, P. 2000. Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, 13(1), p. 403-430.
- Hutchinson, T. & Waters, A. 1987. *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.

- Laufer, B. 1989. What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.). *Special Language: from human thinking to thinking machines* (p. 316-323). Clevedon (England): Multilingual Matters Ltd.
- Lei, L. & Liu, D. 2016. A new medical academic word list: a corpus based study with enhanced methodology. *Journal of English for academic purposes*, 22, p. 42-53.
- Leki, I. & Carson, J. 1994. "Completely different worlds": EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, 31(1), p. 39-69.
- Li, E. & Pemberton, R. 1994. An investigation of students' knowledge of academic and subtechnical vocabulary. In *Joint seminar on corpus linguistics and lexicography* (p. 183-196). Hong Kong: HKUST Language Center.
- Lynn, R. 1973. Preparing word lists: a suggested method. *RELC Journal*, 4(1), p. 25-32.
- McCarthy, M., O'Keeffe, A., & Walsh, S. (2009). *Vocabulary Matrix: understanding, learning and teaching*. Andover: Heinle, Cengage Learning.
- Nagy, W. & Townsend, D. 2012. Words as tools: learning academic vocabulary as language acquisition. *Reading research quarterly*, 47(1), 91-108.
- Nation, P. & Heatley, A. 2002. *Range: A program for the analysis of vocabulary in texts*. Retrieved from <<http://www.victoria.ac.nz/lals/resources/range>>.
- Nation, P. 2013. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesi, H. 2008. *Corpora and EAP*. In: LSP: Interfacing Language with other Realms: Proceedings of the 6th Languages for Specific Purposes International Seminar, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
- Newman, J. A. 2016. *A Corpus-Based Comparison of the Academic Word List and the Academic Vocabulary List*.
- Olsson, E. 2015. Progress in English academic vocabulary use in writing among CLIL and non-CLIL students in Sweden. *Moderna språk*, 109(2), p. 51-74.
- Paquot, M. 2010. *Academic vocabulary in learner writing: from extraction to analysis*. London & New-York: Continuum.
- Park, J. 2012. *Needs analysis of ELI 721 (Undergraduate)*. University of Hawaii at Manoa.
- Praninskas, J. 1972. *American University Word List*. London: Longman.
- Richards, J. 1974. Word lists: Problems and prospects. *RELC Journal*, 5(2), p. 69-84.
- Santos, T. 1988. Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22(1), p. 69-90.
- Shaw, P. 1991. Science research students' composing processes. *English for specific purposes*, 10, p. 189-206.
- Silva, L. G. 2017. Compilation of a Brazilian academic written English corpus. *Revista e-escrita: Revista do Curso de Letras da UNIABEU*, 8(2), p. 32-47.
- Silva, L. G., Matte, M. L., & Sarmiento, S. (2018) Brazilian students' use of English academic vocabulary: an exploratory study. *Linguística de Corpus*, p. 509-526.
- Snow, C. E., Lawrence, J. & White, C. 2009. Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), p. 325-344.
- Snow, C. 2010. Academic Language and the Challenge of Reading for Learning About Science. *Science, New Series*, 328(5977), p. 450-452.
- Stahl, S & Fairbanks, M. 1986. The Effects of Vocabulary Instruction: A Model-Based Meta-Analysis. *Review of Educational Research*, 56(1), p. 72-110.

- Thorndike, E. & Lorge, I. 1944. *The teacher's word book of 30,000 words*. New York: Teachers College Press.
- Townsend, D., Filippini, A., 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), p. 497-518.
- Townsend, D. & Kiernan, D. 2015. Selecting academic vocabulary words worth teaching. *The reading teacher*, 69(1), p. 113-118.
- Trimble, L. 1985. *English for science and technology. a discourse approach*. Cambridge: Cambridge University Press.
- Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. 2009. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28, p. 33-41.
- Wang, J., Liang, S. & Ge, G. 2008. Establishment of a medical academic word list. *English for Specific Purposes*, 27, p. 442-458.
- Ward, J. 2009. A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28, p. 170-182.
- West, M. 1953. *A general service list of English words*. London: Longman.
- Yang, M. 2015. A nursing academic word list. *English for specific purposes*, 37, p. 27-38.

Submetido: 20/06/2018
Aceito: 23/07/2018