Examining the Effectiveness of the LEAD Dictionary in Presenting and Explaining the Top AVL Words

Kazi Amzad Hossain

Lecturer, Department of English, University of Asia Pacific, Dhaka, Bangladesh

ABSTRACT

This study examines the growing importance of proper language usage in academic settings, particularly in English language-based programs, and the shift towards online dictionaries, such as the *Louvain English for Academic Purposes Dictionary* (LEAD), among learners. The LEAD dictionary is lauded for its focus on addressing challenges faced by non-native English writers, including register variability and phraseological patterns. Concurrently, the *Academic Vocabulary List* (AVL), developed by Gardner and Davies, underscores the need for representative academic language in dictionaries. The research evaluates the LEAD dictionary's effectiveness in presenting and exploring the most frequent AVL words, focusing on the top five words: study, group, system, social, and provide. Analysis reveals discrepancies between the presentation of these words in the LEAD dictionary and their frequency and collocations in corpora such as the *British National Corpus* (BNC) and the *Corpus of Contemporary American English* (COCA). The COCA's broader representation of academic disciplines enhances its utility compared to the BNC. Despite being designated as a 'semi-bilingual dictionary', the limited coverage of LEAD for mother tongues restricts accessibility for non-native learners. Suggestions for improvement include expanding language coverage, incorporating additional corpora data, enhancing user-friendly features, providing error feedback, and offering diverse learning exercises. Integration into broader teaching and learning tools warrants further exploration. Overall, the study highlights the discrepancy between the LEAD dictionary's representation of AVL words and their actual usage, signaling areas for enhancement to better serve the needs of language learners and educators.

Keywords: LEAD dictionary, Academic Vocabulary List (AVL), COCA, BNC, Frequency of words.

INTRODUCTION

Using language appropriately is obligatory for academic success. The necessity of adopting proper language in the academic setting is growing along with the increase in English language based study programs. Learners, especially the non-native speakers of English, have been users of traditional "one-size-fits-all" (Rundell 2007: 50) printed dictionaries before the availability of the internet; however, as the internet has become affordable and widespread, learners are leaning towards the online dictionaries for their academic needs. The *Louvain English for Academic Purposes Dictionary* (LEAD) is an online based dictionary which Granger and Paquot (2015:02) claim it to be a "production-oriented online tool" for the non-native learners of English. As they believe that the "register variability" (Granger and Paquot 2010: 88) is a challenge to overcome for both the native and the non-native writers of English, and as, non-native writers face far more difficulties due to cultural and contextual differences, the LEAD dictionary can help overcome these challenges with 'lexical bundles' (Biber et al 1999, quoted in Paquot 2015: 121) among other things. The LEAD also takes into account the learners' usage of phraseological patterns, which are marked by improper collocations and the impact of their native tongues (cf. Granger and Paquot 2010: 89), and assists users in expanding their academic vocabularies (cf. Granger and Paquot 2010: 93). On the other hand, Gardner and

Davies (2014: 306) emphasized that, "[…] it is crucial that the words in any academic list be truly representative of contemporary academic language […]", thus they come up with the Academic Vocabulary List (AVL) derived from an extensive and broad corpus of academic English spanning a wide range of significant academic fields (cf. Gardner and Davies 2014: 312). The AVL includes about 14% of academic content in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) (cf. Gardner and Davies 2014: 305).

The current study evaluates the effectiveness of the LEAD dictionary in presenting and exploring the most frequent AVL words. I will analyze the top five (5) AVL words – *study* (n), *group* (n), *system* (n), *social* (adj) and *provide* (verb), to see how these are presented in the LEAD, and how the frequency and the collocations vary in the BNC and in the COCA corpus. I will also look into the functionality of the LEAD website, and the scope of improvement, if any.

LITERATURE REVIEW

The *Louvain English for Academic Purposes Dictionary* (LEAD) was created by Sylviane Granger and Magali Paquot to address the general academic requirements of the students (cf. Granger and Paquot, 2010; Granger and Paquot, 2015). They described the overall layout of the LEAD dictionary to be "corpus-based, production-oriented, hybrid and customizable" (Granger and Paquot, 2015: 118). It is a web based *English for Academic Purposes* (EAP) dictionary tailored for non-native writers, and, designed to function as an integrated tool, with its dictionary section linked to other language resources. As a corpus-based tool, it is built on the analysis of about 900 academic words and phrases encompassing the academic segment of the British National Corpus (BNC), along with various discipline specific corpora created in-house. Additionally, the dictionary's extensive coverage is enhanced by English as a Foreign Language (EFL) learner corpora from diverse L1 population (cf. Granger and Paquot, 2010: 90 – 94). Granger and Paquot (2010) also believes that, the LEAD dictionary is distinct for its "customizability" (p. 87), although "the degree of customization of the LEAD is admittedly still quite modest: only part of the content is customizable and the tool qualifies as 'adaptable' rather than 'adaptive'" (Granger and Paquot (2015:05). Furthermore, "corpora available in the LEAD is lemmatized and part-of-speech tagged" (Granger and Paquot, 2015: 134).

As Corson (1997) observes, "Good knowledge of academic vocabulary is essential for success at higher levels of education" (quoted in Coxhead 2000: 230), a number of word lists or vocabulary lists, such as the Academic Word List (AWL), have been developed in the recent years. Webb and Nation (2017: 18) believes, "The best-known list of academic vocabulary is the Academic Word List" by Coxhead (2000). They also note,

The Academic Vocabulary list (AVL) (Gardner & Davies, 2014) was developed as a possible successor to the AWL. It provides a comprehensive database of over 3,000 academic words that are ranked according to their frequency in academic discourse, making it a valuable resource of academic vocabulary (2017: 19 – 20).

While considering the essential qualities of an academic vocabulary list, Gardner & Davies (2014: 312) agree that, "The new list must initially be determined by using lemmas, not word families." The word families within the AWL led to numerous issues related to meaning, primarily because they did not account for grammatical parts of speech such as nouns, verbs, adjectives, and adverbs (cf. Gardner & Davies (2014: 307). On the other hand, the *Corpus of Contemporary American English* (COCA) "the 120-million-word academic corpus was already tagged for grammatical parts of speech (e.g. nouns, verbs, adjective, adverbs) by the CLAWS 7 tagger from Lancaster University" (Gardner & Davies (2014: 313). According to Davies (2010: 447), "The Corpus of Contemporary American English is the first large, genre-balanced corpus of

any language, which has been designed and constructed from the ground up as a 'monitor corpus' […]". Monitor corpora are dynamic because new texts continue to be added, unlike the 'static' corpus such as the British National Corpus (BNC) which remains unchanged after its creation (cf. Davies 2010: 447). It is to be noted that the British National Corpus (BNC) contains texts from the 1980s to 1993, while the COCA contains words from the years 1990 – 2019[1]. As the AVL is based on the COCA, and the LEAD is built on the BNC, LEAD seems to be in an 'outdated' position with its words from the BNC until 1993. In the next section, I will describe the methodology that I followed to look into the effectiveness of the LEAD dictionary with regards to the AVL words.

METHODOLOGY

In this research, I will use the AVL data in the LEAD dictionary, as well as in the BNC corpus. As already mentioned, the top five AVL words according to Gardner & Davies (cf. 2014: 317) are – (1) *study* (n), (2) *group* (n), (3) *system* (n), (4) *social* (adj), and (5) *provide* (v). First, I will check out whether these words are included in the LEAD dictionary, and, if yes, then, how are the associated words or collocations being displayed in the LEAD compared to the BNC. Secondly, I will inquire these words and their frequencies in the academic section of the BNC, and compare them with the academic section of the COCA. As the size of the COCA is considerably bigger than the BNC, the frequency of the words will perhaps be reasonably lower in the BNC, thus suggesting the LEAD to be a not-so-contemporary source of vocabulary usage information. Finally, I will inspect the LEAD website to test the functionality of the site, and the possibility of improvement.

DATA ANALYSIS AND RESULTS

The LEAD dictionary landing page has the option to select from 12 disciplines, 5 mother tongue backgrounds, and if, a user's mother language is different from those five, then there is an option to choose 'other' as a mother tongue from the drop down menu. One can also select a country from the drop down menu[2]. Hence, I selected 'General EAP' as discipline, 'other' as mother tongue, and 'Bangladesh' as country to continue to the next web page. The reason for choosing *General EAP* or general *English for Academic Purposes* is that, firstly, it contains 501 texts which is more than the other disciplines in LEAD, secondly, as I will be focusing on the AVL words, I need similar types of words, i.e. academic words in another database. In the next page of LEAD, I searched the first word of the AVL – *study (n)*, however, the search result only showed *study* as a verb.

Figure 1: Search result of 'study' in the LEAD dictionary

It only presented two dictionary meanings and one example in-context each. On the right hand side (as in Fig. 1), there are three sections such as – collocations, lexical bundles, and an option to get more examples from the general EAP discipline. As I clicked on the *General EAP* option, it showed examples from 501 texts, however, these examples have the inflected forms of the word *study*, including nouns. So, it seems unusual that the LEAD dictionary only showed *study* as a verb, even though it has also been used as noun, in the corpus examples at the general EAP option. Besides, although the proposed word suggestions in the collocations and the lexical bundles sections (Fig. 1) are clickable or hyperlinked, some words do not have any entries. For example, I clicked on the word *certain* which is mentioned 9 times in the collocations box, however, none of these *certain* has any entries (Fig. 2 below).
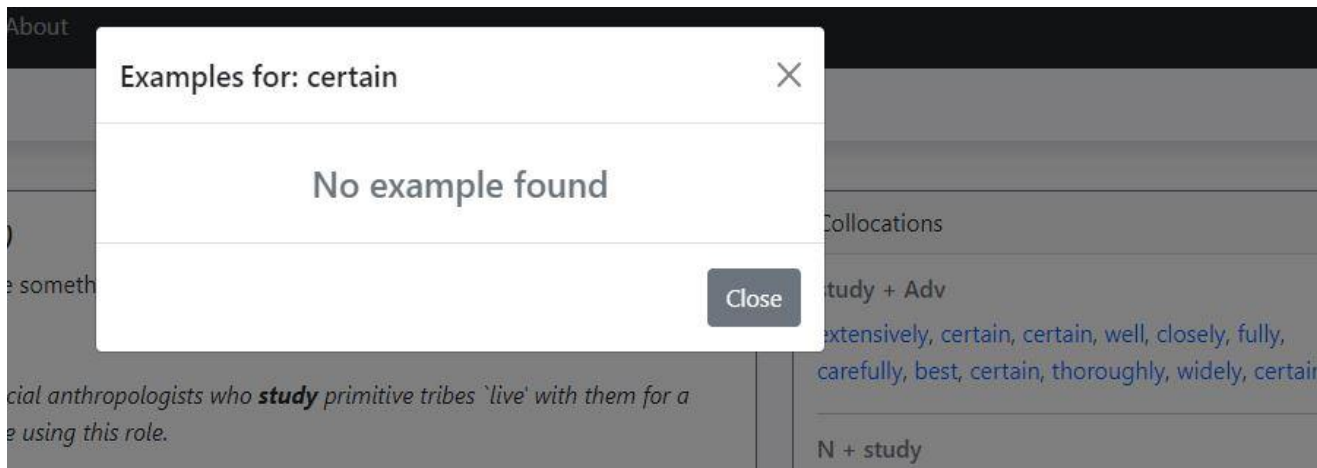

Figure 2: Click result of *certain* in the collocations box in the LEAD website.

Although, all the other clickable word associations showed examples in a pop-up window, it is deviant that *certain* has no entries, which can be deemed a drawback of the LEAD.

The second word of the AVL in discussion is *group (n)*, which, unlike the first word, *study* is listed in the LEAD as the same part of speech as in the AVL. In the LEAD, the word *group (n)* has two meanings with one example each:

group (n.)

1. a set of people, animals, or things that are considered together because they are connected in some way, or share similar features

Despite this decline, the efficacy of the vaccine remained high, across all age **groups**.

2. a set of people who meet or do something together

Associations and pressure **groups** are trying to block the project.

(https://leaddico.uclouvain.be/search/word/group)

However, some hyperlinks in the collocations and in the lexical bundles box remain invalid after being clicked. Besides, general EAP results of the word *group* includes both noun and verb forms.

The third word is *system (n)*, and it is available in the LEAD dictionary after being searched. It also has two meanings with examples, however, some associated words in the collocations box of the website does not work. For example: there are no example or entry records of the words *criminal + system* (Adj +system) as
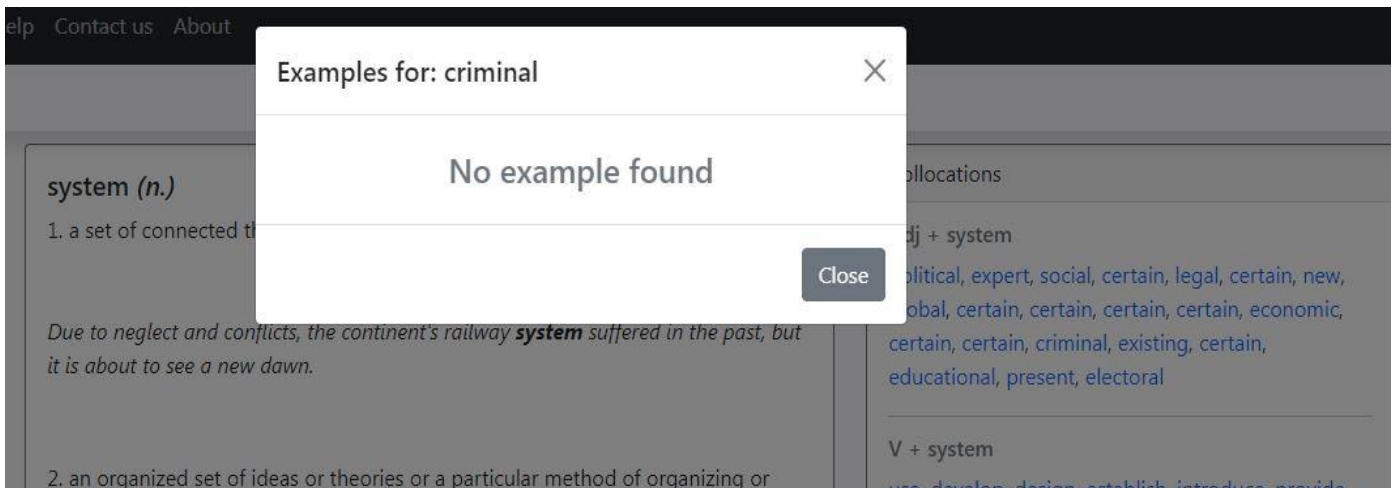
in figure 3 below.

Figure 3: Entry record of *criminal* in the collocations box in LEAD dictionary.

Similar to the previous AVL words, *system (n)* also has more contextual examples with its inflected forms in the LEAD. Those are visible once clicked on 'General EAP' section of the web page.

The fourth word of the AVL is *social (adj.)*. However, it is interesting to note that the word does not exist in the LEAD dictionary. While searching for the word *social (adj.)*, the web page suggested *socially (adv.)* as in figure 4 below shows.
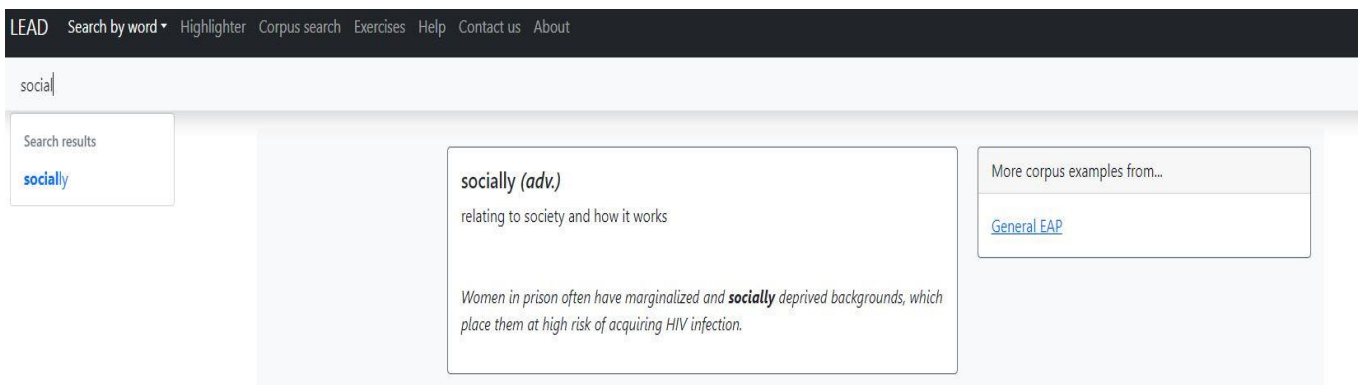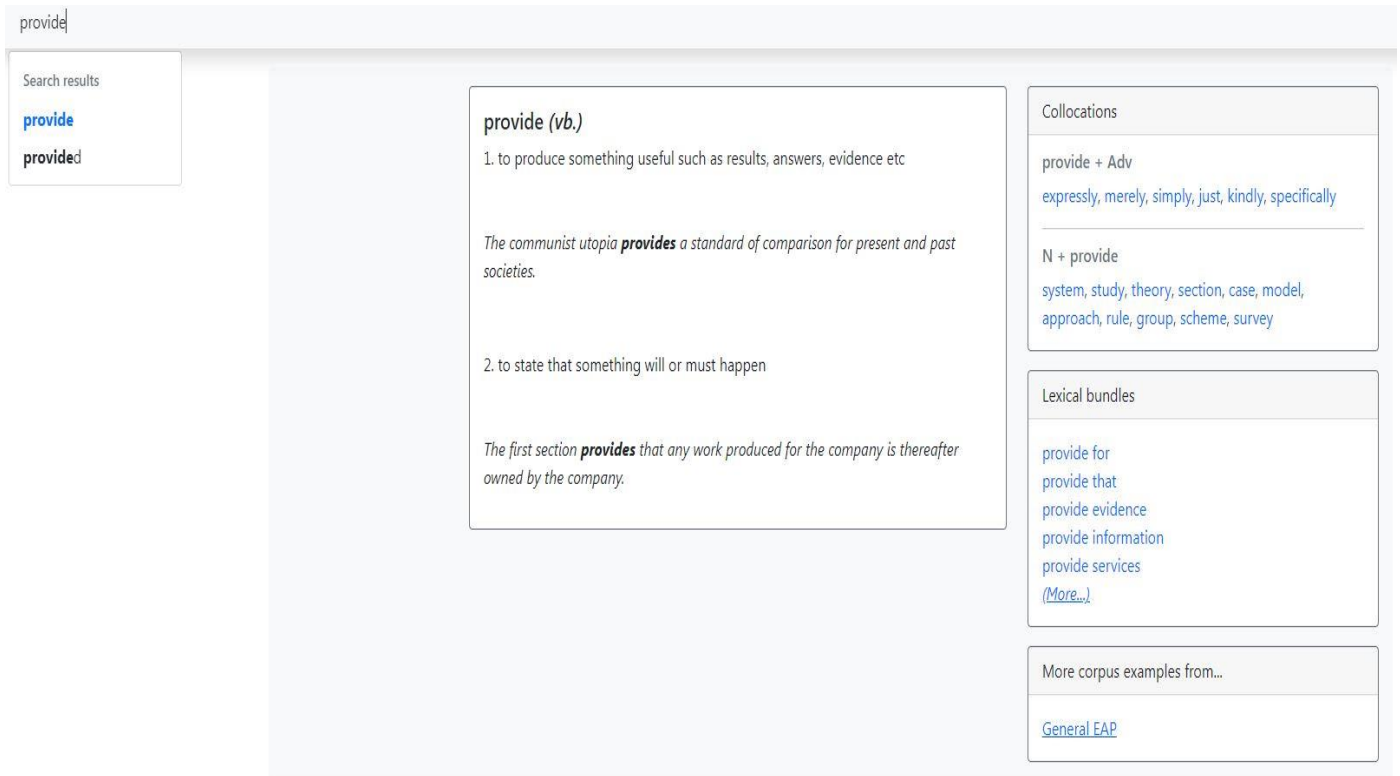

Figure 4: search suggestion of *socially* instead of the keyword *social* in the LEAD dictionary.

The web page also shows that, the entry of *socially* (figure 4) does not have any collocations or lexical bundles suggestions in the LEAD dictionary, which can be regarded as a shortcoming of this online dictionary.
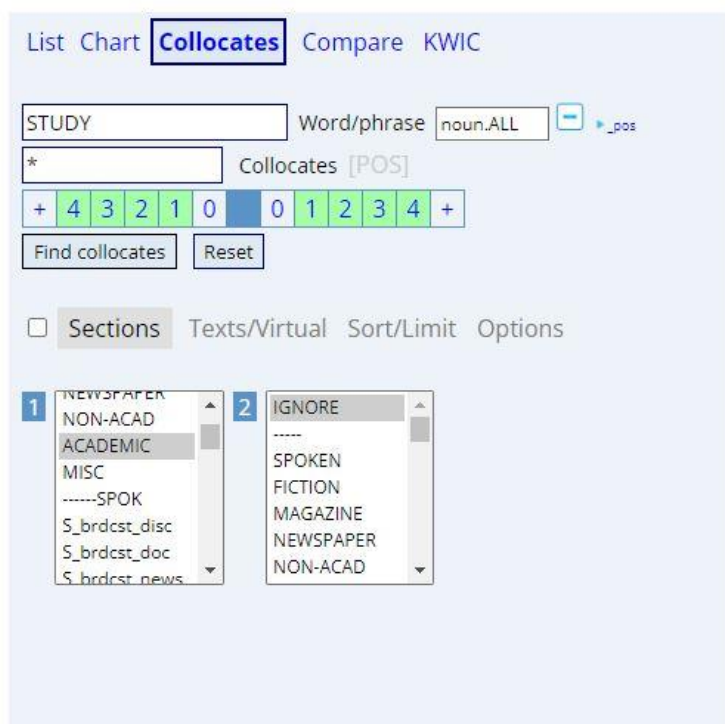
The fifth and final word of my discussion is *provide (verb)*. It has an entry in the LEAD with two meanings and one example each. There are multiple examples of collocations and lexical bundles for this keyword. However, one inconsistency seems to be the suggestion of the inflected form *provided* (figure 5 below) while typing in the word *provide*. With the keywords *study (n)* and *group (n)* there was no suggestions of inflected forms while typing in the search box. It can be considered as a flaw in the LEAD dictionary's organizational pattern for the high frequency words.

Figure 5: search suggestions of the word *provide (v)* in the LEAD dictionary.

Now, I will look into the BNC corpus to see the collocations of these five (5) AVL words. To begin with, I searched STUDY as lemma, as in all nouns, in the academic section (as in Figure 6 below) to find out the most frequently associated words with it.

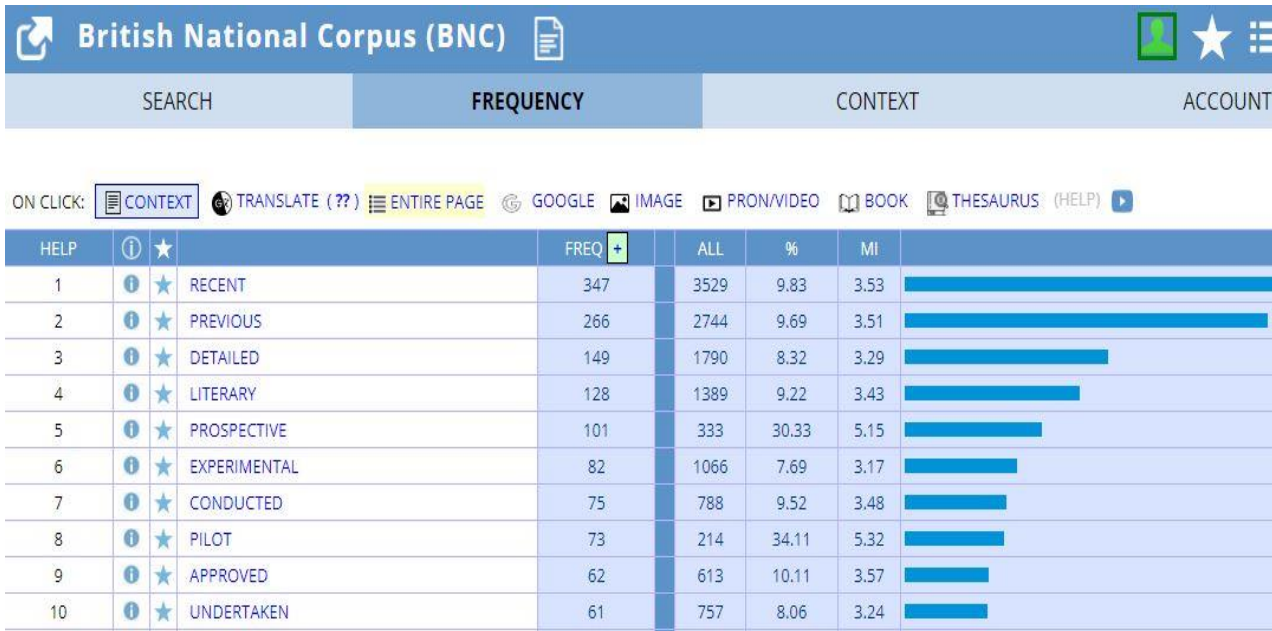Figure 6: *Collocates* search of the lemma *study* in the BNC.

Figure 7: Most frequent words associated with *study* as a noun in the BNC.

Figure 7 shows the top frequent words associated with the word *study (noun)* in the BNC. However, none of these words appeared in the top collocations in the LEAD, perhaps because LEAD does not have *study* as a noun.

Figure 8 below shows the top results of the collocations of *group (n)* in the BNC academic section. The LEAD also does not represent that. For example, in the LEAD, the top three lexical bundles are – 'a group of', 'control group', 'group of people'[3], while in the BNC, are – 'age group', 'individuals group', and 'pressure group' as in figure 8.
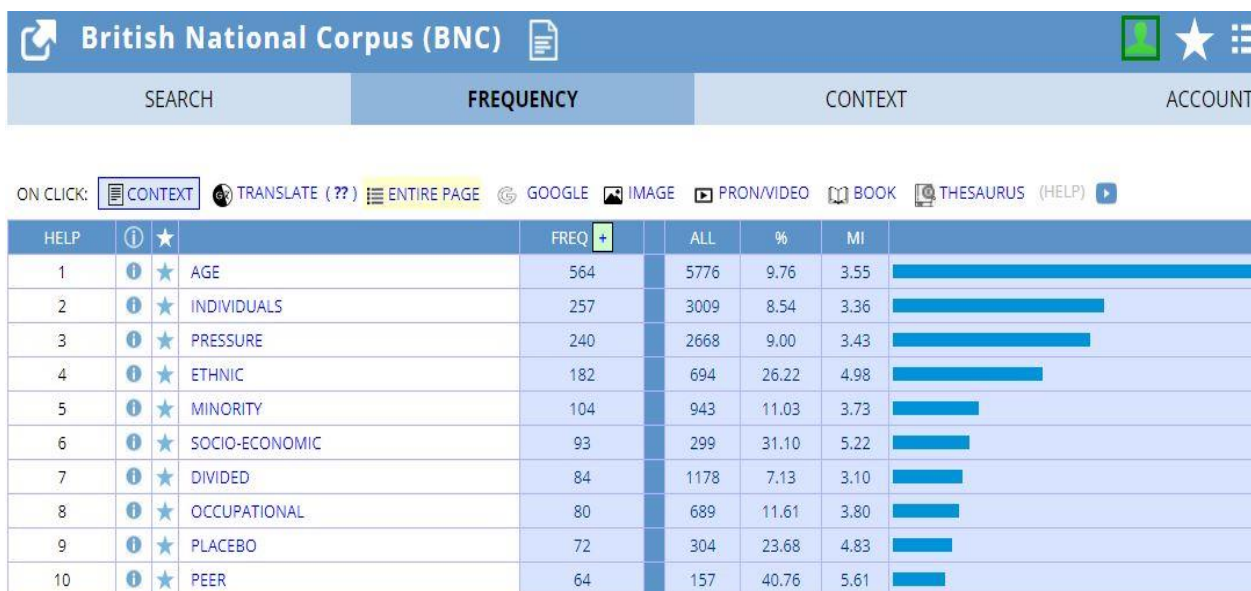
Figure 8: Most frequent words associated with *group* as a noun in the BNC.

The *collocates* search of *system (n)* as a lemma in the BNC produces the following figure (figure 9). The top collocates words with *system* are – 'nervous', 'expert', and 'recognition' in the BNC, however, there are no examples of *system + noun* collocations in the LEAD dictionary.
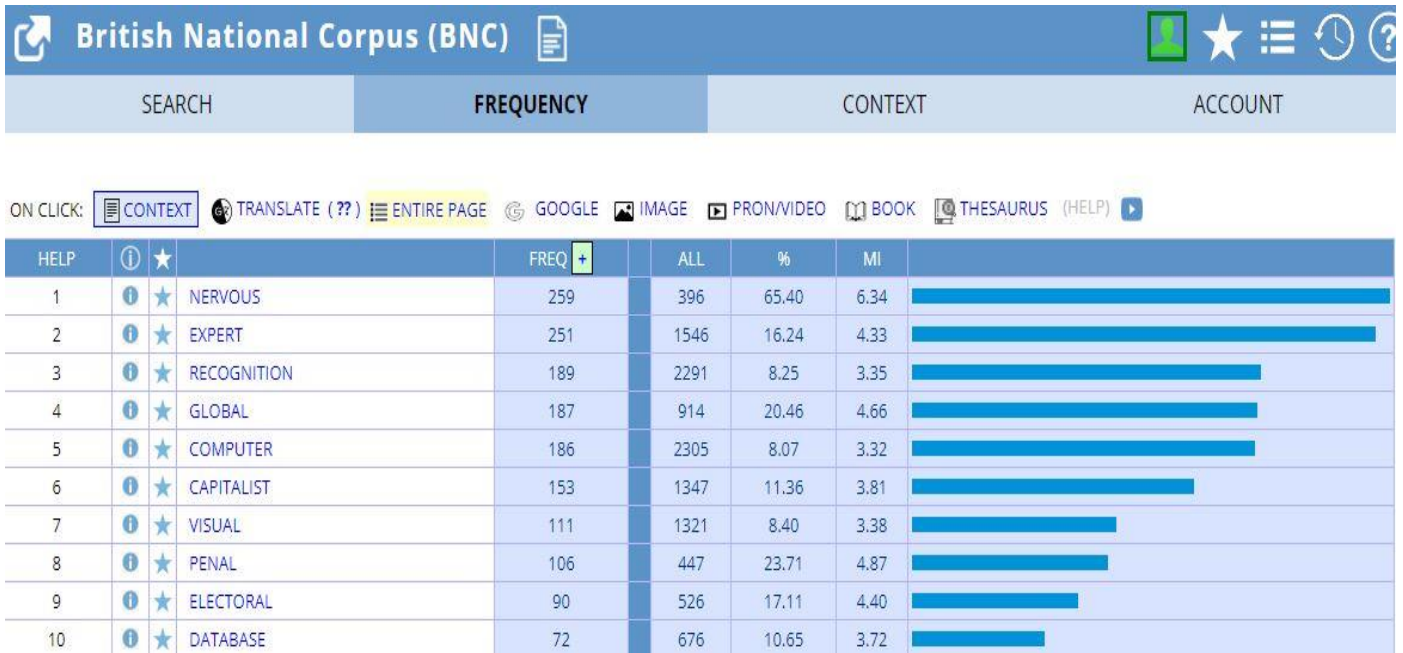
Figure 9: Most frequent words associated with *system* as a noun in the BNC.

Next, I searched *social (adj.)* as lemma in the academic section of BNC as in figure 10 below.

Figure 10. Collocates search of the lemma *social (adj.)* in the BNC.

Figure 11 below illustrates the three most frequent words in the BNC with *social (adj.)* are – 'services', 'economic', and 'class'. In contrast, figure 4 clearly depicts that, *social (adj.)* does not occur in the LEAD dictionary. So, the fourth most frequent word of AVL is non-existent in the LEAD.

Figure 11. Most frequent words associated with *social* as an adjective in the BNC.

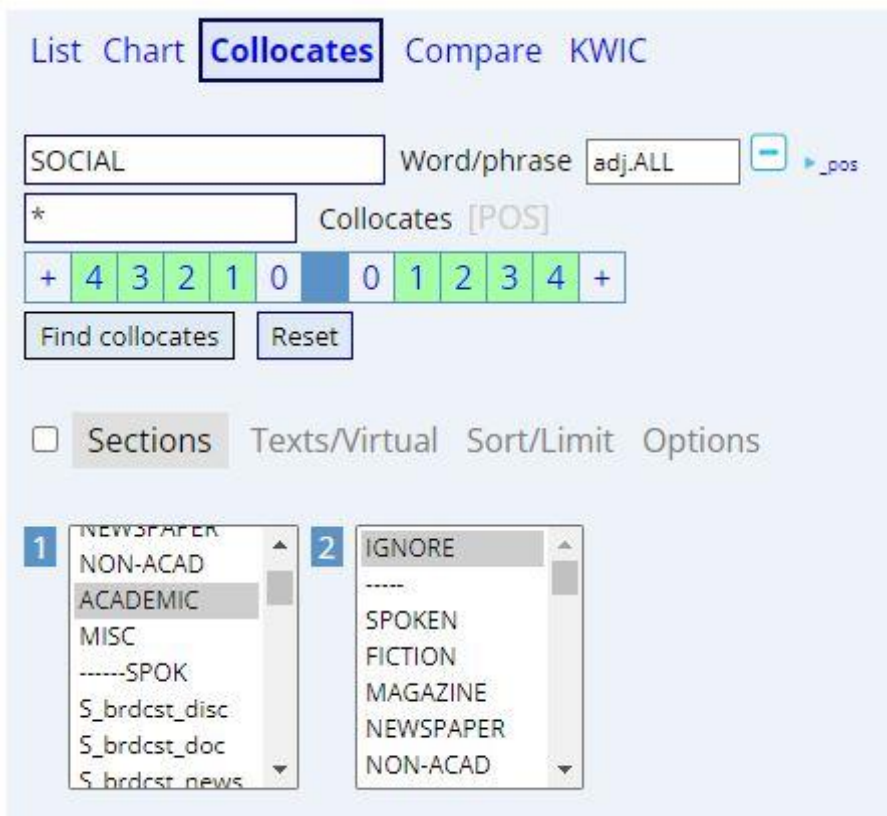Finally, I searched *provide (verb)* as lemma similarly in the academic section of BNC. The top three words associated with it are – 'services', 'useful', and 'framework' as shown in figure 12 below.

Figure 12. Most frequent words associated with *provide* as a verb in the BNC.

It can be observed in the LEAD dictionary that, the lexical bundles 'provide services', and 'provide a framework for' do exist. Out of the five words, only *provide (verb)* has two matches in the LEAD. In the *lexical bundles* of the LEAD it has seventeen hyperlinked phrases.

Now I will look into the comparative frequency of these five AVL words in the academic section of the BNC and in the COCA corpus. To do this, I will do a list search of all these words (lemma) separately in both the BNC and in the COCA corpus to find the total frequency, then I will only consider the AVL word form, not other inflected forms of the word. After compiling the results I will present them in a table below (table 1).

No.	AVL words	Parts of Speech tag	Frequency		Total Frequency	
			BNC	COCA	BNC	COCA
01	STUDY	noun.ALL	8898	133786	16345	219306
02	GROUP	noun.ALL	8820	85424	15941	148474
03	SYSTEM	noun.ALL	11452	75783	15501	111493
04	SOCIAL	adj.ALL	17846	124418	17846	124418
05	PROVIDE	verb.ALL	5419	50895	13109	118225

Table 1: The frequency of the top five AVL words in the academic sections of the BNC and in the COCA corpus.

For the purpose of better visualization, I have produced a chart below (chart 1) from these data.
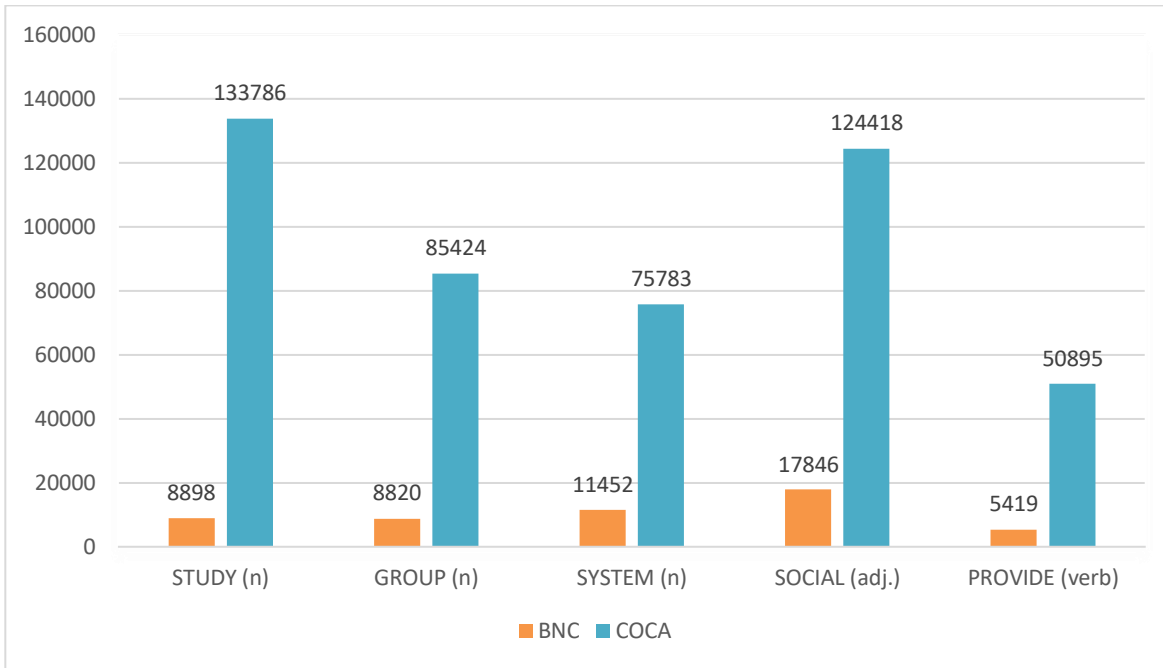
Chart 1: The comparative frequency data of BNC and COCA from table 1.

Before considering these data in the chart, we have to remember the size difference of the BNC and the COCA as I discussed in the literature review of this study. BNC has data of roughly 13 years until 1993, while COCA consists data of 30 years until 2019.

From chart 1 it can be noticed that, compared to the COCA, all the AVL words have very low frequency in the BNC. As LEAD dictionary is established from the BNC data, it can be deduced from this chart that, LEAD does not cover the most used academic vocabulary or collocations of the recent times, while COCA is highly relevant in comparing collocations. Even if the motivation behind developing the LEAD was to help non-native learners, one might ask nowadays, why a non-native learner of English would be drawn to the LEAD dictionary, while there are more recent dictionaries like the *Oxford English Dictionary*, or, databases like the COCA which provides a great quantity of examples and collocations for the high frequency academic words. Besides, corpora like COCA have far more useful functionality than the LEAD website, for example, various search categories such as *list*, *chart* etc.; options to manipulate data by selecting and deselecting *sections*, *texts* buttons etc.; and evaluating a great range of contextual sentences.

While using the LEAD dictionary website, I have come to realize that there are certain things that are not working, or not updated in the website. For example, after clicking on the *highlighter* button, which was supposed to highlight errors and/ or suggest dictionary entries[4] upon input of texts, a new page is displayed, however, no matter what text I inserted, and then clicked on *process* button, it did not work. It showed an error message as in figure 13 below.
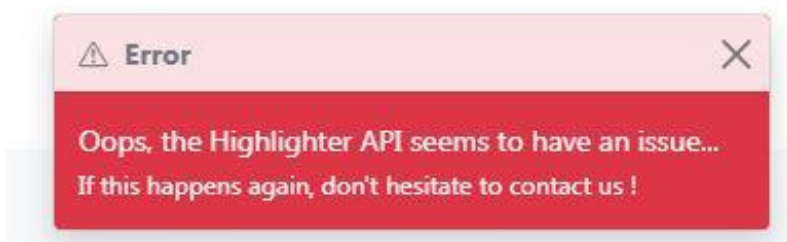

Figure 13: Error notice in LEAD dictionary while using the *highlighter* option.

Similar errors occurred when I tried to test the *thumbs up* or *down* button of a hyperlinked word as in figure 14 below. Once an example is given (clicked on) a *thumbs up* or *down*, it cannot be undone. Clicking multiple times to remove the *up* or *down* vote generates this error message. It is not clear what function does these buttons play – whether it impacts the example search results or is it just for decoration purposes with green and red colors. There are no specific instructions on the *help* page of the LEAD website.
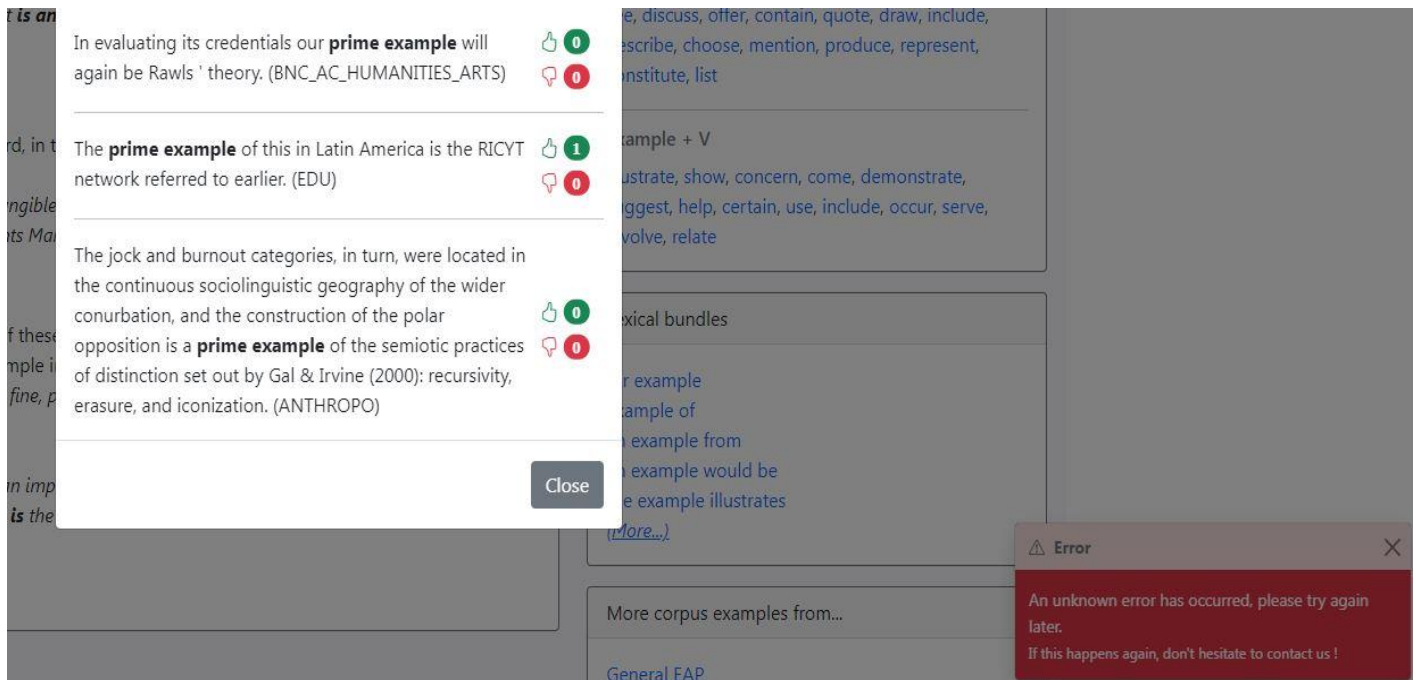

Figure 14: Error note after clicking on *thumbs up/ down* sign multiple times.

Five mother tongue backgrounds, namely – French, Dutch, Spanish, Chinese and German, and one *other* option are available in the LEAD dictionary, which seems inadequate for the vast majority of the non-native speakers of English who lives in other countries. Besides, the drop-down menu of the mother tongues is not alphabetically organized.

DISCUSSION

In this section, I would like to discuss the results of the data analysis of the previous section. It is worth noticing that out of five AVL words, two, i.e. *study (n)* and *social (adj.)* did not occur in the LEAD dictionary. Other words even if was present, nonetheless the collocations or the word associations was not abundant enough. The word *group (n)* has only one match in the LEAD dictionary, e.g. *pressure group*, compared to the BNC, however, it is not in the top three examples of lexical bundles in the LEAD dictionary. The word *system (n)* in the LEAD has also only one matching example with the top BNC hits, e.g. *expert + system*. The word *provide (verb)* in the LEAD has two matches with the BNC hits, however, these are not in the top three examples of the *lexical bundles* in the LEAD. The frequency distribution in table 1 earlier, shows that there is a tremendous difference between the BNC and the COCA in representing the *Academic Vocabulary List* data. In this regard, it can be mentioned that, in the *academic* category of BNC, there are six sub-disciplines, namely – law, medicine, humanities, natural science, social science and engineering. COCA has however ten sub-disciplines under the category *academic*, namely – history, education, geography/ social science, law/ political science, humanities, philosophy/ religion, science/ technology, medicine, business and miscellaneous. This distribution also makes the COCA superior to the BNC. Another drawback of the BNC is that it does not have the detailed word or lemma search option like in the COCA as in figure 15 below.

Figure 15: Detailed information for a word or lemma search window of the COCA.

After a lemma search in the *word* category of the COCA, the top five AVL words' ranks in the overall corpus can be found as in table 2 below.

Words	STUDY (n)	GROUP (n)	SYSTEM (n)	SOCIAL (adj)	PROVIDE (v)
Rank	259	199	219	344	272

Table 2: rankings of the top five AVL words according to COCA corpus.

The ranks show that these words are significant enough to be in a frequent word list, and Gardner and Davies (2014) appropriately put them in the *Academic Vocabulary List* (AVL). However, from the data analysis and the overall findings, it is evident that, the LEAD dictionary could not represent the highly frequent AVL words properly, and the explanation of the AVL words in the LEAD lacks suitable demonstrative examples.

Although LEAD is called a "semi-bilingual dictionary" (Laufer and Levitzky-Aviad 2006, quoted in Granger & Paquot 2010: 91), it has only five mother tongues listed, therefore, it would not be as useful to English non-native learners of most countries except those five. Perhaps adding some more native languages would make it relevant to the learners of more countries. Also, adding more data from other representative corpora can help LEAD be a better reference tool of comparison. Although Granger & Paquot (2015:12) reports that, "errors and difficulties found in the writings of a wide range of learner populations are dealt with in generic notes […]", this feature is not noticeable for the top AVL words. Another weak point of LEAD is that, sometimes learners does not realize that they made some mistakes, in that case, LEAD cannot help if they do not have "a certain degree of language proficiency for effective corpus consultation" (Chang 2014: 255). One scope of improvement for the LEAD dictionary is perhaps to get rid of all the error messages that occurs while searching for a word and make the site more user friendly. Besides, adding more exercises for various levels of non-native learners would make it attractive to a new audience. Addition of audio-visual data can also draw traffic into the website. The LEAD web architecture can also be improved for better navigability of the users. Although the authors Granger & Paquot have pointed out that the LEAD may be "integrated into a general dictionary and/ or a suite of teaching and learning tools" (2010:95), there is no further evaluation in sight.

CONCLUSION

In this study, I used the top five AVL words to find out whether these are adequately represented in the

LEAD dictionary. In the literature review I described the development of the LEAD dictionary and the Academic Vocabulary List. I also mentioned the years covered of the BNC and the COCA. In the methodology section, I referred to the tasks to be done such as looking for word frequencies in the LEAD, and compare those with the BNC and the COCA corpus. In the fourth section, I described the data collection processes from the LEAD, the BNC and the COCA, and provided necessary illustrations and examples to prove that the representation of the AVL words in the LEAD dictionary is very poor. The lemma search results of the AVL words in the BNC and in the COCA reinforces that the LEAD is not efficient enough to demonstrate the AVL words. In the discussion section, I compared the results, looked into some examples, and compiled the ranks of the AVL words in the COCA corpus. I also put forward some recommendations to improve the LEAD dictionary based on my evaluation.

REFERENCES

1. Biber, Douglas., et al. 1999. *Longman Grammar of Spoken and Written English*. Edinburgh: Pearson Education Ltd.
2. Chang, Ji-Yeon. 2014. The use of general and specialized corpora as reference sources for academic English writing: A case study. Re*CALL* 26(2), 243-259.
3. Corson, David. 1997. The Learning and Use of Academic English Words. Language Learning, 47 (4): 671-718. https://doi.org/10.1111/0023-8333.00025
4. Coxhead, Averil. 2000. A New Academic Word List. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951
5. Davies, Mark. 2004. *British National Corpus* (from Oxford University Press). Available online at https://www.english-corpora.org/bnc/.
6. Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.
7. Davies, Mark. 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. In Literary and Linguistic Computing (Vol. 25, Issue 4, pp. 447–464). Oxford University Press (OUP). https://doi.org/10.1093/llc/fqq018
8. Gardner, Dee, & Davies, Mark. 2014. A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015
9. Granger, Sylviane & Paquot, Magali. 2010. *Customising a general EAP dictionary to meet learner needs. Proceedings of ELEX2009.*In: Granger S.; Paquot M., *ELexicography in the 21st century: New Challenges, new applications*, Presses universitaires de Louvain : Louvain-la-Neuve2010, p.87-96. http://hdl.handle.net/2078.1/75727
10. Granger, Sylviane & Paquot, Magali. 2015. Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid / Die elektronische Lexikographie wird spezifischer: Das Design und die Struktur einer auf die Benutzerbedürfnisse berzogenen akademischen Online-Schreibhilfe / La lexicographie électronique devient plus spécifique: conception et structure d'une aide à l'écriture académique. *Lexicographica*, *31*(1), 118-141. https://doi.org/10.1515/lexi-2015-0007
11. Laufer, Batia & Levitzky-Aviad, Tamar. 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' – A Dictionary for Production in a Foreign Language. International Journal of Lexicography. 19. 10.1093/ijl/eck006.
12. LEAD: The Louvain EAP Dictionary. https://leaddico.uclouvain.be/ (06 December 2023).
13. Rundell, Michael. 2007. The dictionary of the future. In Granger, S. (ed.) Optimising the Role of Language in Technology-enhanced Learning. Proceedings of Expert Workshop of the Integrated Digital Language Learning seed grant project, Louvain-la-neuve, Belgium: 49-51.
14. Webb, Stuart, & Nation, Paul. 2017. *How Vocabulary is Learned*. Oxford University Press. ISBN: 978 0 19 440352 8 eBook.

FOOTNOTES

[1] https://www.english-corpora.org/bnc/

[2] https://leaddico.uclouvain.be/

[3] https://leaddico.uclouvain.be/search/word/group

[4] https://leaddico.uclouvain.be/help#highlighter