CORPUS-BASED APPROACH TO VOCABULARY TEACHING AND LEARNING
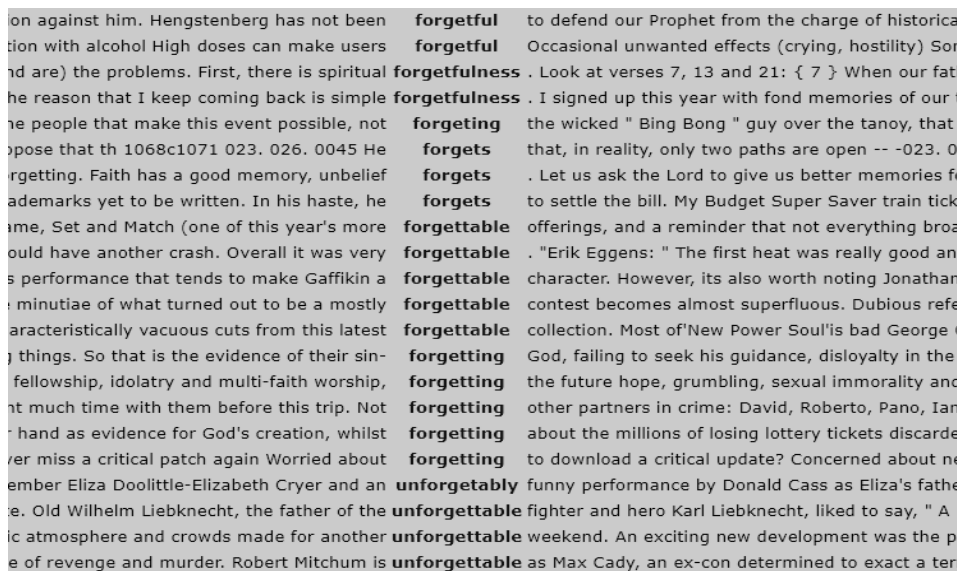
Asst. Prof. Dr. Emin Idrizi
International Balkan University
eminidrizi@gmail.com
North Macedonia

Abstract. University EFL learners are typically required to learn a vast amount of vocabulary in a short period of time in order to cope with challenging English tasks, such as comprehending academic texts or writing assignments. One effective strategy that can be used to boost learners' vocabulary is the word form strategy (Cook, 2016). However, a valid question that arises is: how should word forms be taught? One traditional and widely applicable method is the deductive approach, which typically implements the PPP (Presentation-Practice-Production) strategy. The invention of corpora and their applicability in language education, however, have paved the way for a new approach to language learning. The corpus-based approach, which is seen as an inductive approach and that normally adopts the III (Illustration – Interaction - Induction) strategy, could be used as an alternative approach to teaching word forms. In this paper, the author discusses some reasons why vocabulary expansion through corpora deserves consideration in language pedagogy.

Keywords: Vocabulary learning, Word forms, Data-driven learning, Corpora, Learner autonomy

1. INTRODUCTION

Since the invention and the advancement of electronic corpora, new perspectives to language investigation have emerged. Most importantly, corpora have become accessible to everyone who has access to computers and advanced language analyses are now easier and faster than ever before. For instance, information on some of the most common inflections and derivations, i.e. word family, of the word *forget* in the corpus (*Figure 1*) is available at the touch of a button and investigations of this kind can be carried out for any other word we wish to explore.

Figure 1 An excerpt from Skylight web interface showing some common inflections and derivations of the word 'forget'

This is only a simple query that we can launch in a corpus as many other advanced language enquiries can be made. For instance, corpora are capable of investigating a word's common phrases and patterning, collocation, frequency, context, to name a few. What we can discover from a corpus investigation may be often hard, if not impossible, to find in advanced grammar books or dictionaries. The high potential of modern time corpora in language study, as a result, has sparked considerable interest in their role in the language education. Corpus-based teaching and learning or Data-driven learning (DDL) thus has become a field of extensive research and it has drawn the attention of many theorists as well as language practitioners.

2. CORPORA AND VOCABULARY LEARNING

Various studies have been conducted in the last decades to investigate the efficacy of corpus-based approach to vocabulary teaching and learning (see Cobb, 1999; Yılmaz and Soruç, 2015; Ashkan & Seyyedrezaei, 2016). They largely indicate positive results, thus seeing DDL as a valuable learning approach. More importantly, these studies, as well as other studies carried out in the field of DDL (Gaskell and Cobb, 2004; Yoon & Hirvela, 2004; Gilmore, 2009; Mull, 2013), generally show that learners have positive attitudes towards using corpora as reference and learning tools.

Theorists and practitioners point to numerous benefits learners obtain when learning vocabulary through corpora. Referring to several researchers and authors, Nation (2001) summarizes several advantages of corpus-based vocabulary learning:

1. Learners meet vocabulary in real contexts. The information which these provide often differs from non-corpus-based descriptions.
2. Multiple contexts provide rich information on a variety of aspects of knowing a word, including collocates, grammatical patterns, word family members, related meanings and homonyms.
3. The use of concordances involves discovery learning, where the learners are being challenged to actively construct generalizations and note patterns and exceptions.
4. Learners control their learning and learn investigative strategies. (p. 111)

Wilson (2013) holds that corpus-based pedagogy could do more than what dictionaries and textbooks do in language investigation and language learning, e.g. in the expansion of the vocabulary. For Quan (2016), corpus consultation "may provide focused repetitions of the target word, as learners are offered the opportunity to go through a number of examples in a short time, which may take years for them to meet via conventional reading" (p. 276). Similarly, Gilquin and Granger (2010) maintain that extensive encounter of say certain words can lead to the development of learners' vocabulary as well as the phraseologies with which they are associated.

3. LEARNING WORD FORMS THROUGH CORPORA

Word forms have drawn the attention of many theorists and practitioners and have been viewed as a useful strategy to boost language learners' vocabulary. Cook (2016), for instance, maintains that making students aware of the fact that words with certain affixes have associated meanings may be beneficial in two ways. Firstly, learners may infer the meaning of a word based on the knowledge of affixes attached to that word. Secondly, they can use the strategy efficiently to memorize considerable vocabulary. That is to say, by having learners know the meaning of the prefix *hyper* – meaning more than normal or too much – they would be able to both comprehend and memorize with no trouble a considerable number of other words with this sort of word formation, such as *hyper*active, *hyper*creative, *hyper*sensitive, and so on.

Word forms have been traditionally taught through common deductive approach. That is to say they follow the prominent PPP (Present-Practice-Produce) strategy: (1) word forms are first presented to learners; (2) then they are given time to practice them; (3) finally, learners are expected to produce the same. The deductive approach is a top-bottom approach in which teaching and learning moves from rules towards specific examples and it is traditionally seen as a more teacher-centered approach.

With the advancement of electronic corpora and the development of DDL as an alternative approach, however, a new method to teaching and learning word forms could be said to have emerged. Below we discuss how word forms could be taught to language learners through corpus-based approach.

As opposed to the traditional pedagogy, Data-driven learning, as the name implies, is an inductive or a bottom-up approach in which teaching and learning moves from examples to making generalizations or rules. As a result, data-driven learning can adopt the III strategy (McEnery & Xiao, 2011) which stands for Illustration, Interaction, and Induction. In the illustration stage, learners look at concordance lines; in the interaction stage, they share their findings from corpus with their peers or teacher; lastly, they make generalizations or infer rules based on their findings from the corpus data. In the case of teaching and learning word forms with the help of corpora, the following stages may be included: learners may be firstly required to investigate the most common words that contain a certain affix (e.g. *hyper-* mentioned earlier). Additionally, they may be tasked to find meanings or grammatical class of the words using concordance lines for clues. In the second phase, they could share their findings with their peers or teachers. Finally, in the deduction phase, learners are supposed to make generalizations based on the findings from the corpus data. It is clearly evident that DDL is a more student-centered approach in which learners are seen as "Sherlock Holmes" (Johns, 1997, p. 101) or language investigators rather than simple recipients of knowledge. Teachers, on the other hand, are no more seen as 'teachers' in the traditional sense, but rather as 'facilitators' in the language classroom (Warren, 2016).
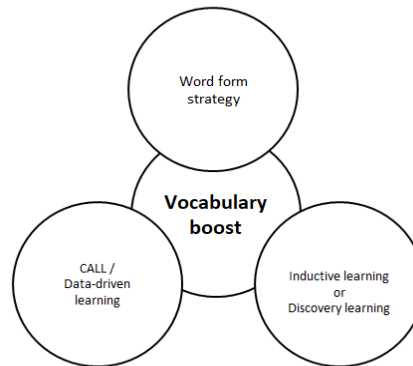
Figure 2 An illustration of corpus-based vocabulary learning in combination with word form strategy

There are numerous corpora online that could be used to investigate words with affixes. One corpus that is designed to be more user-friendly for research, teaching, and learning is *IntelliText* (Sharoff, 2014) which is administered by University of Leeds. As pointed out in one of its tutorials, this corpus has a special function which could be used to search and study words based on affixes and it is considered very useful when it comes to boosting learners' vocabulary. Searches can be done using various English text corpuses, such as British National Corpus (BNC), UK webpages corpus (UKWAC) and so on, depending on what variety or type of text one wants to investigate. In the case of word forms, BNC would be more appropriate since it is considered to represent general English. *IntelliText* has a special feature which searches for words with certain affixes. The searches can be launched by entering either a prefix or suffix in the search box. The results are typically given in a list based on their frequency of occurrence, the most frequent ones appearing first. By clicking the frequency number of one of the words, a learner is able to get the concordance lines for that particular word (Figure 3). The data then could be used for variety of purposes,

such as guessing the meaning of a word, notice grammatical class, find common patterning etc., depending on the task given by the instructor.

prev	1	2	3	4	5	6	7	8	9	10		Concordances for [lemma="criticism"]		
Lemma	Forms	Count ▾		titleid	left	match	right
criticism	6	5735		A03	apologise for challenging public	criticisms	of northern Malawians made by
mechanism	6	4800		A03	imprisoned and tortured because of his	criticisms	of the military government, was
capitalism	4	1844		A03	has moved to try to counter the	criticisms	of its human rights record. In
organism	5	1791		A03	Lankan Government is responsive to	criticism	from outside the country. Unfortunately
socialism	5	1638		A04	SETH. NEEDLES IN HAY: IS IT ART	CRITICISM	? TWO QUESTIONS What is art? Leo
racialism	6	1110		A04	philosophers and aestheticians. Is it	criticism	? There is a massive amount of
nationalism	4	1001		A04	immediately be identified by a reader as	criticism	. Writing by the art critic of
tourism	3	975		A04	a newspaper is self-evidently	criticism	, in parallel with the writing
communism	3	929		A04	are less certainly described as	criticism	, for their main topics may be
realism	4	878		A04	Books and catalogues may contain	criticism	; but their writers may think of
optimism	3	835					
Marxism	2	724					
feminism	4	665					
scepticism	5	644					
terrorism	3	632					

prev 1 2 3 4 5 6 7 8 9 10   next > last >>

Figure 3 The most frequent -isms (list on the left) in English and some concordance lines of the word 'criticism' (examples on the right) investigated in IntelliText using BNC corpus

4. SOME BENEFITS AND DRAWBACKS OF CORPUS-BASED TEACHING AND LEARNING OF WORD FORMS

Learning word forms through corpus-based approach may benefit the language learners in many ways. One very important advantage of learning with corpora is the fact that learners do not rely on invented language, as it is typically the case in many language textbooks, but rather they are exposed to naturally occurring language as originally spoken or written by native speakers of English. This, as a result, may improve motivation since learners are aware that "they are learning the language they will encounter when they step outside the language classroom and into the world of language use" (Ruppen, 2010, p. xi). This suggests that corpus could play a role of a 'linguistic expert' for the learner, whenever he has a question about the language he is learning.

Investigating word forms through corpus consultation can guarantee rich information about words that are often unlikely to be found in grammar books or advanced dictionaries. Learners, by the help of corpus advanced features, may find useful lists of words that contain certain affix ordered by frequency (see Figure 3). This helps them discover the words that are more common in the language (and thus have more priority in vocabulary learning); and vice versa, it helps them avoid words (although this still depends on the learner's proficiency and his circumstances) which occur very rare in English and which they may never encounter or use in their professional or daily lives. They can also extract useful and in-depth information about words they are learning as concordance tools provide plentiful context for words as well as advanced queries for this purpose. Last but not least, learners can extract useful contextual information about words. For instance, when one investigates the concordance lines for context clues about the word *hypertension* notices that it is typically used in medicine.

DDL provides a feeling of empowerment in learners (Gilquin and Granger, 2010). Namely, having learners explore word forms in the corpus can boost their confidence in a way that they too can find useful information about vocabulary just as lexicographers do. In addition, DDL is believed to foster learner autonomy (Bernardini, 2004). This in fact may be one of the most useful aspects of learning word forms through DDL. As there are many word forms in English, it would be impossible to cover them all in the

class. For this reason, it is useful to use the approach many times in the classroom until the learners have become accustomed to the approach. However, it is on learners' own responsibility to continue using this strategy autonomously, beyond college, for the purpose of expanding their vocabulary even further.

Corpus-based learning involves useful cognitive processes. O'Sullivan (2007), for instance, mentions seventeen mental processes taking place when learners engage in corpus investigations. Some of these processes are: deducing, making hypotheses, predicting, and so on. Based on this, it could be assumed that DDL benefits the learner not only linguistically but also with valuable extralinguistic skills that are useful both at college and beyond.

If one decides to use a corpus-based approach to teaching word forms, however, some potential problems should be taken into consideration. One obstacle we may encounter is that a purely inductive approach may be too difficult for some learners, particularly for low proficiency ones or those who are not comfortable with the inductive learning style (see Flowerdew 2009, Huang, 2008). However, DDL is not always seen as a purely inductive approach. For Gilquin and Granger, (2010), for example, corpus-based pedagogy can be adapted anywhere between fully teacher-centered and learner-centered scale. Huang (2008) uses the term 'guided' DDL to advocate a softer version of corpus-based approach (p. 20). What is clear from this discussion is that it is on the instructor to decide whether to provide assistance during his corpus-based class tasks as this will depend on his students' ability to cope with the task. Moreover, if the instructor decides to aid students due to the presence of weak students or those who can't cope with purely inductive teaching and learning, he should also decide on the degree of the assistance. The assistance can be offered by using clues (Flowerdew, 2008) to make sure the approach is made suitable for all learners.

Another obstacle we could encounter while using corpora to teach word forms is that of learners coping with the software. Corpus interfaces have become more and more user-friendly during the last decades but not to a degree to say that learners can handle it without preparation. According to Gilquin and Granger (2010), corpus training is necessary if one decides to use DDL in the classroom. This is due to lack of preparation, both in handling the software functions as well as in interpreting the corpus data, which will most likely result in unsuccessful completion of tasks. Therefore, sufficient preliminary corpus training is a must if we want to have an effective and productive corpus-based pedagogy.

5. CONCLUSION

Based on the discussion in this paper, it is reasonable to believe that learners could effectively use corpora as a reference tool to explore word forms as the new technology is now sufficiently advanced for this purpose. This can be done with a minimum teacher help, i.e. with more learner autonomy, or with a greater degree of support depending on learners' capacities, proficiency, and learning styles. In addition, learners should undergo a sufficient training in the use of corpus software as well as in the skills necessary to interpret concordance lines as these may undermine the effectiveness of the approach. The discussion also leads us to one of the most fundamental questions related to the effectiveness of DDL in vocabulary learning: does corpus-based learning of word forms result in effective retention? If yes, then, does learning word forms through DDL result in better retention compared to the traditional learning approaches? This is an area that is open to more research which would further clarify the effects of teaching and learning of vocabulary through corpora.

REFERENCES

Ashkan, L., & Seyyedrezaei, S. H. (2016). The Effect of Corpus-Based Language Teaching on Iranian EFL Learners' Vocabulary Learning and Retention. International Journal of English Linguistics, 6(4), 190.

Bernardini, S. (2004). Corpora in the classroom. In J. Sinclair (Ed.), How to use corpora in language teaching. Amsterdam and Philadelphia: John Benjamins, 15-36.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. Computer Assisted Language Learning, 12(4), 345-360.

Cook, V. (2016). Second language learning and language teaching (5th ed.). New York, NY: Routledge.

Flowerdew, L. (2008). Corpus-based analyses of the problem-solution pattern: a phraseological approach (Vol. 29). John Benjamins Publishing.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy. International Journal of Corpus Linguistics, 14(3), 393-417.

Gaskell, D., Cobb, T. (2004). Can learners use concordance feedback for writing errors? System, 32(3), 301-319.

Gilmore, A. (2009). Using online corpora to develop students' writing skills. ELT Journal, 63(4), 363–372

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe and M. McCarthy (Eds.), The Routledge handbook of corpus linguistics, 359370.

Huang, L. S. (2008). Using Guided, Corpus-Aided Discovery to Generate Active Learning. In English teaching forum (Vol. 46, No. 4, pp. 20-27). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs, SA-5, 2200 C Street NW 4th Floor, Washington, DC 20037.

Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. Teaching and language corpora, 100-115.

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), Handbook of research in second language teaching and learning, 2, 364-380.

Mull, J. (2013). The learner as researcher: Student concordancing and error correction. Studies in Self-Access Learning Journal, 4(1), 43-55.

Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge: CUP.

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. ReCALL, 19, 269-286

Quan, Z. (2016). Introducing "mobile DDL (data-driven learning)" for vocabulary learning: an experiment for academic English. Journal of Computers in Education, 3(3), 273-287.

Reppen, R. (2010). Using corpora in the language classroom. Cambridge University Press.

Sharoff, S. (2014). IntelliText Corpus Queries [Computer Software]. Retrieved September 9, 2017, from: http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#

Warren, M. (2016). Introduction to data-driven learning. In F. Farr and L. Murray (Eds.), The Routledge handbook of language learning and technology. New York, NY: Routledge.

Wilson, J. (2013). Technology, pedagogy and promotion. The Higher Education Academy.

Yılmaz, E., & Soruç, A. (2015). The use of Concordance for Teaching Vocabulary: A Data-driven Learning Approach. Procedia-Social and Behavioral Sciences, 191, 2626-2630.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. Journal of second language writing, 13(4), 257-283.