

THE DEVELOPMENT OF WRITING COMPETENCE THROUGH DATA-DRIVEN LEARNING (ON THE BASIS OF COCA)

Radjabova Gulnoza

Uzbekistan State World Languages University

ABSTRACT

The purpose of this article is to demonstrate how teachers can use data-driven learning methods to teach writing. First, a brief overview of corpus linguistics, data-driven learning, and the corpus used in this article (COCA) is given. Data-driven learning has been proved as an effective approach in helping learners solve various writing problems such as correcting lexical or grammatical errors, improving the use of collocations and generating ideas in writing, etc. This article reports on an empirical study in which data-driven learning was accomplished with the assistance of the user-friendly Corpus of Contemporary American English (COCA), and presents some strategies of how this corpus can be implemented by English teachers in the classroom with special focus on writing.

Keywords: Data-driven learning (DDL), EFL learners, writing development, corpus COCA, Corpus Linguistics;

INTRODUCTION

Corpus linguistics is empirical as it thrives on data to analyze and discover what language speakers do. Large electronically stored texts (corpora) reveal these patterns in grammar, and discourse. Corpora allow access to authentic data revealing knowledge about the language that is not intuitive and show frequency patterns of words and grammar constructions. Such patterns can be used to improve language materials or to directly teach learners. As J. Flowerdew (2015) claims: “no dictionary or grammar is able to fully describe the language” (p. 329) so corpus linguistics fills in the gaps describing missing pieces that students can hardly learn traditionally from other data.

With the advent of faster and more powerful computers in the eighties, along with other turns in linguistic theory, corpus linguistics has turned into a viable linguistic methodology. Since that time, it has only continued to grow in use and applications because it can be applied to every branch of linguistics (McEnery, Wilson, 2001). Therefore, the most recent branch of linguistics to benefit from applications of corpus linguistics is language teaching. According to Romer (2009), “corpus linguistics can make a difference for language learning and teaching having an immense potential to improve pedagogical practice” (p. 84). Bennett (2010) describes three ways corpora can be incorporated into language teaching: corpus-influenced materials, corpus-cited texts, and corpus-designed activities. Corpus-influenced materials include textbooks and classroom materials that are not just based on traditional grammar rules, but patterns and frequency information obtained from corpora. Corpus-cited texts are dictionaries and grammar books based on corpus data. Corpus-designed activities are those in which the learner actually sees or is involved with the actual data in some way, or essentially DDL.

DDL is a term coined by Tim Johns (1991) to describe language learners acting as “language detectives.” According to him, it is a way of using corpus linguistics in teaching by exposing learners to data and having them find the rules and patterns from example concordance lines, or lines of text from a corpus that focus around a single word, which means that “the language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data”, therefore, the term “data-driven learning” (DDL) has been implemented to describe the approach (Johns, 1991, p. 2).

There are some benefits of DDL implementation and the authentic language of DDL is one of the greatest benefits that is cited in its favor (Romer, 2008). In this way, DDL allows teachers and students to study naturally occurring language for grammatical patterns, word usage, semantic and pragmatic features, and textual discourse (L. Flowerdew, 2009). Additionally, authentic language allows for language topics that otherwise might not occur in a textbook or from the examples created in teachers’ minds, to be used in vocabulary and grammar discussions. Finally, since teachers cannot change the language students encounter in English speaking environments, DDL, in providing authentic examples and encouraging “noticing” helping students of all levels develop skills needed to deal with the language around them (Romer, 2008). Another possible benefit of DDL is that it allows for more potential learner autonomy in the classroom, with less reliance on the teacher. This autonomy, being based on inductive and deductive reasoning methods, gives opportunities for the development of cognitive skills (Boulton, 2009b).

In order to provide a useful tool for DDL, an appropriate corpus must be selected, particularly one that enables to develop writing skills and is accessible enough for DDL. All activities in this article are based on The Corpus of Contemporary American English (COCA) (Davies, 2008). Principle reasons for using COCA are its authenticity, dynamism (with 20 million words being added on a yearly basis), large size, diverse structure, user-friendly interface, easy accessibility and ability to show features of vocabulary which is very important in teaching writing.

Foreign/second language writing usually poses great challenges to English as foreign language (EFL) learners, whose writing are usually regarded as “non-native-like”. The writing problems, such as lexical poverty and miscollocations, are commonly seen in their writing. DDL, a corpus-assisted language learning approach, is considered as advantageous in solving learners’ writing problems. But to ensure the effectiveness of DDL activities, a careful design should be considered. An examination of previous DDL studies in writing demonstrates that the following factors may affect the learning outcomes of DDL activities.

The first important factor is task type since not all writing problems can be solved equally well by DDL approach. In previous studies, DDL is adopted to learn linking adverbials (Cotos 2014), to aid learners in generating ideas and writing creatively (Kennedy and Miceli 2010), to distinguish synonym adjectives and solve learners’ problems of overusing general adjectives to learn collocation knowledge and thus help learners produce more accurate and complex language patterns (Thomas 2015). The second significant factor is the methodology, that is, how DDL is implemented. There are generally two ways to adopt DDL approach, namely, direct DDL and indirect DDL (Yoon and Jo 2014). In direct DDL, learners consult corpora directly for solving language problems; whereas indirect DDL refers to learners’ use of paper-

based concordance lines extracted by teachers or researchers. The third factor to be considered is language proficiency and training. It is usually believed that DDL may appeal more to advanced learners. Just as shown by O'Sullivan and Chambers (2006), the post-graduates showed more positive attitudes toward corpus use than the undergraduates. What's more, Granath (2009) also stated that advanced learners usually benefit more from a DDL approach. Nonetheless, with proper training, DDL could be also effective for lower-level learners as well as advanced learners. As stated by Boulton (2009), for intermediate or lower-level learners, it is crucial to organize the training session to improve their corpus techniques for DDL to be successful. Thus, it is clear that task type, methodology, language proficiency and training are important variables to be considered while designing DDL activities to develop writing. It is also obvious that the carefully designed DDL activities are effective in helping learners solve some writing problems.

May Fan (2009) studied the use of English collocations by ESL students and found that these students used fewer collocations than native speakers and were unable to collocate words correctly. She also found interference from their native language collocations. According to her, "collocational use is not rule-governed and, in most cases, arbitrary and idiosyncratic, it is important for teachers to raise the awareness of L2 learners to this problematic aspect of language to develop their writing skills" (Fan, 2009, p. 115). Therefore, corpora help identify collocates, so DDL can be an effective way to increase students' collocational awareness to improve their writing skills.

There are several different ways to see how a word or its collocates vary in frequency across register. The most basic way is to choose the "chart" radio button at the top of the screen and enter the word or phrase in the "word" query box. This will show the distribution of the word across registers as well as time. To get more specific sub-register information simply click on the register title above the bar graphs. To see KWIC or example sentences for each sub-register or collocations it is necessary to click on this bar in COCA. It is also possible to perform basic frequency and collocate searches and specify what register or specific sub-register to search in. The default is set at "ignore" but settings can be specified to search for a word with a KWIC or collocate display from a particular sub-register. Learners can choose more than one sub-register or collocation as well by holding control on the keyboard while selecting. As an example of choosing certain collocations and registers, students writing an academic paper might want to see example sentences of using key words only in academic writing and therefore choose only the academic register.

Thus, in this article we investigated the effects of DDL activities on EFL learners' writing development in terms of collocability and register. The results confirmed the effectiveness of adopting DDL in writing essays and developing learners' writing accuracy in terms of collocations and correct vocabulary choice.

REFERENCES

1. Bennett, G. R. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Ann Arbor: University of Michigan Press.
2. Boulton A (2009) Testing the limits of data-driven learning: language proficiency and training. *ReCALL* 21(1):37-54

3. Cotos E (2014) Enhancing writing pedagogy with learner corpus data. *ReCALL* 26(2):202–224
4. Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. Available online at <http://www.americancorpus.org>
5. Flowerdew, J. (2009). Corpora in language teaching. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 327-350). Oxford, UK: Wiley-Blackwell.
6. Johns, T. (1991). "Should you be persuaded": two samples of data-driven learning materials. *English Language Research Journal*, 4, 1-16.
7. Kennedy C, Miceli T (2010) Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Lang Learn Technol* 14(1):28–44
8. Fan, M. (2009). An exploratory study of collocational use by ESL students – A task-based approach. *System*, 37, 110-123. doi: 10.1016/j.system.2008.06.004
9. McEnery, T., Wilson, A. (2001). *Corpus Linguistics* (2nd ed.): Edinburgh University Press.
10. Romer, U. (2008). Corpora and language teaching. In A. Ludeling & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 112-130). Berlin: Mouton de Gruyter.
11. O'Sullivan Í, Chambers A (2006) Learners' writing skills in French: corpus consultation and learner evaluation. 15(1):49–68
12. Thomas, J. (2015). Stealing a march on collocation. *Mult Afford Lang Corpora Data-driven Learn* 69:85–108
13. Yoon H, Jo JW (2014) Direct and indirect access to corpora: an exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Lang Learn Technol* 18(1):96–117.