Spectrum Journal of Innovation, Reforms and Development
Volume 49, March 2026 | ISSN (E): 2751-1731
Website: www.sjird.journalspark.org

DATA-DRIVEN LEARNING FOR COLLOCATION DEVELOPMENT IN SECONDARY EFL CLASSROOMS: THEORETICAL FOUNDATIONS AND PRACTICAL IMPLICATIONS

Bahronov Nodirbek Aliyevich
Independent Researcher, Department of Interfaculty Foreign Languages,
Bukhara State University
e-mail: bahronovnodirbek@gmail.com

Abstract

Collocations – conventional word combinations such as make a decision or heavy rain – are a key to accurate and natural use of English, yet secondary school learners rarely receive systematic instruction in them. This article reviews the theoretical foundations of data-driven learning (DDL) and examines its suitability for developing collocational competence of upper-grade (grades 8–9) EFL learners. Drawing on research in corpus linguistics, the lexical approach and the noticing hypothesis, the paper identifies the mechanisms through which corpus consultation supports collocation learning and discusses the constraints of using DDL with younger, lower-proficiency learners. It concludes that DDL becomes effective at secondary level when corpus tools are learner-friendly, tasks are guided and closely linked to the textbook, and the teacher acts as a facilitator.

Keywords: Data-driven learning, collocation, corpus, secondary school, EFL, noticing, lexical approach.

Introduction

Over the last three decades, research in corpus linguistics has shown that a large part of natural language consists of recurrent word combinations rather than freely assembled words (Sinclair, 1991; Wray, 2002). Native speakers choose make a mistake rather than do a mistake and heavy traffic rather than strong traffic not because of grammatical rules but because of conventional usage. Pawley and Syder (1983) described this ability as “nativelike selection” and argued that it underlies fluent speech.

For learners of English, collocations are notoriously difficult. Studies of learner writing consistently find that even advanced learners produce fewer and less varied collocations than native speakers and make frequent errors, especially in verb + noun combinations (Nesselhauf, 2003; Laufer & Waldman, 2011). In Uzbekistan, where foreign language education has been declared a national priority (Resolution PQ-5117, 2021), upper-grade school students are expected to move from basic to independent use of English, yet textbooks still present vocabulary mainly as single words.

Data-driven learning (DDL), a term introduced by Johns (1991), offers a possible solution. In DDL learners examine authentic examples from a corpus and discover patterns of use themselves. This article aims to (1) summarise the theoretical foundations of DDL relevant to collocation learning, and (2) identify the conditions under which DDL can be applied in secondary school EFL classrooms.

Spectrum Journal of Innovation, Reforms and Development

Theoretical foundations

Collocation in linguistic theory. The modern notion of collocation goes back to Firth (1957), whose dictum "you shall know a word by the company it keeps" placed word combinations at the centre of meaning. Sinclair (1991) developed this idea into the "idiom principle", according to which speakers rely on a large stock of semi-preconstructed phrases. Corpus methods made it possible to measure such co-occurrence objectively, giving teachers empirical evidence about which combinations are typical.

The lexical approach. Lewis (1993, 2000) argued that language consists of "grammaticalised lexis, not lexicalised grammar" and that collocations should be a central element of the syllabus. He proposed that learners should be trained to notice chunks in input, record them as wholes and recycle them in output. These recommendations are directly compatible with DDL, since corpus concordances present many instances of the same chunk in context.

Noticing. Schmidt's (1990) noticing hypothesis holds that input becomes intake only when learners consciously attend to it. Collocations are particularly prone to being "unnoticed", because learners understand take medicine without difficulty and therefore pay no attention to the verb. Concordance lines, in which the node word is centred and its neighbours are aligned, make such patterns visually salient and thus promote noticing.

Data-driven learning. Johns (1991) described the learner as a researcher who formulates questions about language and answers them by examining corpus evidence. Gilquin and Granger (2010) distinguish between direct (hands-on) consultation of corpora by learners and indirect use of teacher-prepared corpus materials. Boulton (2010) showed that paper-based concordances can be effective even without computers, which is important for schools with limited technology.

Evidence on effectiveness

The effectiveness of DDL has been examined in several meta-analyses. Boulton and Cobb (2017), synthesising 64 studies, found large overall effects of corpus use, with particularly strong effects for vocabulary and lexico-grammar. Lee, Warschauer and Lee (2019) confirmed positive effects of corpus use on vocabulary learning in a multilevel meta-analysis. Focusing specifically on collocations, Vyatkina (2016) found that corpus consultation helped learners of different proficiency levels with collocation tasks, often more than conventional reference resources.

At the same time, most DDL research has been conducted with university students. Boulton (2009) showed that even lower-intermediate learners can benefit from corpus work with little training, provided that the materials are suitable, while Chen and Flowerdew (2018) note that long, complex concordance lines and insufficient training remain obstacles. Crosthwaite (2017) showed that corpus-based feedback is most useful when learners know what to look for. These findings indicate that DDL cannot be transferred to secondary schools without adaptation.

Implications for secondary EFL classrooms

Based on the literature reviewed, four conditions for applying DDL to collocation learning in grades 8–9 can be formulated.

1. Learner-friendly corpus tools. Corpora designed for linguists often have complex query syntax. For school learners, tools such as SkELL (Sketch Engine for Language Learning), which displays short, readable example sentences and "word sketches" grouped by grammatical relation, are more suitable (Kilgarriff et al., 2015). The larger Corpus of Contemporary American English (COCA) can be used for frequency and genre comparisons under teacher guidance (Davies, 2008–).

2. Guided discovery tasks. Instead of open-ended searches, learners need structured questions such as "Which verbs come before decision?" or "Which adjectives describe rain?". Frankenberg-Garcia (2014) showed that several carefully selected corpus examples can effectively support learners' language production.

3. Integration with the textbook. Corpus work should not be an isolated activity. When search words are taken from the current textbook unit (for example, meal, dish and order in a unit on eating out), learners see the immediate relevance of the corpus and recycle the vocabulary of the lesson.

4. The teacher as facilitator. In DDL the teacher selects the words, prepares or checks the concordances, monitors learners' conclusions and corrects over-generalisations. Römer (2011) emphasises that teachers themselves need corpus literacy in order to play this role effectively.

Conclusion

Theoretical and empirical research provides strong grounds for using data-driven learning to develop collocational competence. The approach is supported by the Firthian view of meaning, the lexical approach and the noticing hypothesis, and its effectiveness for vocabulary and collocation learning has been confirmed by meta-analytic evidence. For secondary school learners, however, DDL requires learner-friendly tools, guided tasks linked to the textbook, and an active facilitating role of the teacher. These conditions formed the basis of the methodology developed in our research for grade 8–9 learners in general secondary schools of Uzbekistan.

REFERENCES

1. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning, 67(2), 348–393.
2. Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. Language Learning, 60(3), 534–572.
3. Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. ReCALL, 21(1), 37–54.
4. Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. International Journal of Corpus Linguistics, 23(3), 335–369.
5. Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. Computer Assisted Language Learning, 30(6), 447–473.
6. Davies, M. (2008–). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

7. Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In Studies in linguistic analysis (pp. 1–32). Blackwell.
8. Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. ReCALL, 26(2), 128–146.
9. Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), The Routledge handbook of corpus linguistics (pp. 359–370). Routledge.
10. Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. English Language Research Journal, 4, 1–16.
11. Kilgarriff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015). Corpora and language learning with the Sketch Engine and SKELL. Revue française de linguistique appliquée, 20(1), 61–80.
12. Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. Language Learning, 61(2), 647–672.
13. Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. Applied Linguistics, 40(5), 721–753.
14. Lewis, M. (Ed.). (2000). Teaching collocation: Further developments in the lexical approach. Language Teaching Publications.
15. Lewis, M. (1993). The lexical approach: The state of ELT and a way forward. Language Teaching Publications.
16. Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. Applied Linguistics, 24(2), 223–242.
17. Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), Language and communication (pp. 191–226). Longman.
18. Resolution of the President of the Republic of Uzbekistan No. PQ-5117 of 19 May 2021 "On measures to bring the activity of popularising the study of foreign languages to a qualitatively new level". https://lex.uz
19. Römer, U. (2011). Corpus research applications in second language teaching. Annual Review of Applied Linguistics, 31, 205–225.
20. Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158.
21. Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press.
22. Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. Language Learning & Technology, 20(3), 159–179.
23. Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press.