Neo Science Peer Reviewed Journal
Volume 43, February - 2026 ISSN (E): 2949-7701
www.neojournals.com

EFFECTIVENESS OF A CORPUS-BASED METHODOLOGY FOR DEVELOPING COLLOCATIONAL SKILLS IN WRITTEN AND ORAL SPEECH: AN EXPERIMENTAL STUDY IN SECONDARY SCHOOLS OF BUKHARA REGION

Bahronov Nodirbek Aliyevich
Independent Researcher,
Department of Interfaculty Foreign Languages,
Bukhara State University
E-mail: bahronovnodirbek@gmail.com
ORCID: 0009-0006-2228-9548

Abstract:

This article reports the results of an experimental study that tested a corpus-based methodology for developing collocational skills in the written and oral speech of upper-grade learners. The study was carried out in the 2025–2026 academic year in 13 general secondary schools – one in each of the 11 districts and the two cities of Bukhara region – with about 1500 learners of grades 8 and 9. The experimental groups were taught with the "Corpus Detective" exercise system and the "CORPUS-BASED COLLOCATIONS' LAB" website using the SkELL and COCA corpora, while the control groups followed the standard textbook. By the end of the experiment, experimental-group learners used 70–80% of collocations correctly in their written and oral speech, and first-language-based collocation errors became markedly less frequent in their work. The findings suggest that corpus-based instruction adapted to the school context is an effective way of developing collocational skills.

Keywords:

Collocational skills, corpus-based methodology, experiment, SkELL, COCA, written speech, oral speech, secondary school.

Introduction

Collocational knowledge is closely linked to the accuracy and fluency of learners' language (Pawley & Syder, 1983; Wray, 2002). Studies of learner language show that inappropriate word combinations are among the most frequent errors in learners' writing (Nesselhauf, 2003; Laufer & Waldman, 2011). Meta-analyses indicate that corpus-based instruction, known as data-driven learning, is effective for vocabulary and collocation learning (Boulton & Cobb, 2017; Lee et al., 2019). However, most of this evidence comes from university settings, and there is little research on secondary school learners, particularly in Central Asia.

The present study tested a corpus-based methodology developed for grade 8–9 learners in general secondary schools of Uzbekistan. The research question was: How does corpus-based

This work is published under **CC BY-NC-ND 4.0**

instruction affect the correct use of collocations in the written and oral speech of grade 8–9 learners?

Methodology

Participants and setting. The experiment was conducted in the 2025–2026 academic year in 13 general secondary schools of Bukhara region: one school from each of the 11 districts and one each from the cities of Bukhara and Kogon. About 1500 learners of grades 8 and 9 participated. In each school, classes were assigned either to an experimental group (EG) or to a control group (CG). Both groups consisted of learners of the same grades who followed the same curriculum with the same number of English lessons per week.

Intervention. Throughout the 2025–2026 academic year, the EG was taught with the corpus-based methodology. The methodology included the six-stage "Corpus Detective" exercise system (Notice – Search – Analyze – Practice – Produce – Reflect) and tasks on the "CORPUS-BASED COLLOCATIONS' LAB" website. Tasks were based on the textbook units of grade 8 ("Eating out", "Street food around the world", "The latest technology", "Computers and the internet") and grade 9 ("It's a challenge", "Our changing planet", "Cultural understanding: National parks", "Our holiday", "My place", "Strange houses"). Collocations and examples were drawn from SkELL (Kilgarriff et al., 2015) and COCA (Davies, 2008–). In schools with limited internet access, teacher-prepared printed concordances were used (Boulton, 2010). The CG studied the same units with the standard textbook exercises.

Instruments. Collocational skills were measured with a pre-test and a post-test, each consisting of a written and an oral part. The written part included a gap-fill task, an error-correction task and a short essay on a unit topic. The oral part was an individual interview with a two-minute talk on a unit topic. In both the essay and the talk, the rate of correct collocation use was calculated as the number of correctly used collocations divided by the total number of collocations used. The correctness of collocations was verified against SkELL and COCA. Oral performance was also rated on four criteria – accuracy, variety, fluency and appropriacy – on a 0–3 scale (Hughes, 2003).

Stages of the experiment. The experiment was organised in three stages, summarised in Table 1.

Table 1. Stages of the experiment

Stage	Aim	Main activities
Ascertaining (pre-test)	to determine the initial level of learners' collocational skills	written and oral diagnostic tests; analysis of typical collocation errors
Formative	to apply the corpus-based methodology in the EG	"Corpus Detective" tasks on the website; SkELL and COCA searches; speaking and writing tasks based on textbook units
Final (post-test)	to evaluate changes in learners' collocational skills	repeated written and oral tests; comparison of results across groups and stages

This work is published under CC BY-NC-ND 4.0

Data analysis. The rates of correct collocation use were compared across groups and test stages. Qualitative data from learners' essays and spoken tasks were analysed to identify changes in the types of collocations used and in typical errors.

Results

At the ascertaining stage, learners in both groups showed similar patterns of collocation use. In their essays and spoken answers they relied mainly on single words and on combinations translated from Uzbek or Russian, and errors such as *do a mistake*, *drink medicine*, *strong rain* and *make a photo* were common.

At the final stage, the rate of correct collocation use in the written and oral speech of EG learners reached 70–80%. The improvement was observed in both grades and in schools with different levels of technical equipment, including schools where printed concordances were used instead of direct corpus access. A detailed statistical comparison of the EG and CG results will be presented in a separate publication.

Qualitative observation showed that EG learners increasingly used verb + noun collocations from the units (*book a table*, *download an app*, *face a challenge*, *protect the environment*) and corrected typical errors such as *do a mistake* and *strong rain*. In the oral test, EG learners paused less often when using the collocations practised in the Produce stage.

Discussion

The results indicate that corpus-based instruction adapted to the secondary school context is effective for developing collocational skills in both written and oral speech. This is consistent with meta-analytic evidence (Boulton & Cobb, 2017; Lee et al., 2019) and with Vyatkina's (2016) finding that corpus consultation supports collocation learning for learners of different proficiency levels.

Three features of the methodology are likely to explain its effectiveness. First, the Notice and Analyze stages drew learners' attention to collocations as units, in line with the noticing hypothesis (Schmidt, 1990). Second, the Produce stage required learners to use the collocations in their own writing and speaking, which is necessary for productive knowledge (Nation, 2013). Third, the Reflect stage taught learners to check their own collocations against corpus evidence, which made them more independent.

The study has limitations. The intervention lasted one academic year, and the long-term retention of collocations was not measured. Differences in internet access between schools may also have influenced how learners worked with the corpora. Future studies could include a delayed post-test and compare direct and paper-based corpus use.

Conclusion

The experiment conducted in 13 schools of Bukhara region with about 1500 grade 8–9 learners showed that a corpus-based methodology combining the "Corpus Detective" exercise system, the "CORPUS-BASED COLLOCATIONS' LAB" website and the SkELL and COCA corpora improves the correct use of collocations in learners' written and oral speech: by the end of the

This work is published under CC BY-NC-ND 4.0

experiment, learners in the experimental group used 70–80% of collocations correctly. The methodology can be recommended for English lessons in general secondary schools and for teacher training courses.

REFERENCES

1. Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. Language Learning, 60(3), 534–572.
2. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning, 67(2), 348–393.
3. Davies, M. (2008–). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/
4. Hughes, A. (2003). Testing for language teachers (2nd ed.). Cambridge University Press.
5. 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.
6. Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. Language Learning, 61(2), 647–672.
7. 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.
8. Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge University Press.
9. Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. Applied Linguistics, 24(2), 223–242.
10. 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.
11. Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158.
12. Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. Language Learning & Technology, 20(3), 159–179.
13. Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press.

This work is published under **CC BY-NC-ND 4.0**