DATA-DRIVEN LEARNING: ELEVATING LEXICOGRAMMAR AND AUTONOMY IN ENGLISH LANGUAGE ACQUISITION THROUGH CORPUS TOOLS

A'zamova Mohiro'y
English teacher
moxiazamova@gmail.com
DOI:https://doi.org/10.5281/zenodo.23063444

Traditional English language instruction frequently relies on artificial, textbook-driven examples that fail to capture the nuanced realities of authentic communication. This pedagogical limitation often leaves English as a Second/Foreign Language (ESL/EFL) learners struggling with lexical choices, contextual grammar, and authentic phraseology. Over the past few decades, corpus linguistics has emerged as a revolutionary framework offering direct access to vast, empirical databases of naturally occurring language. By utilizing software tools to query these corpora, language pedagogy can shift from deductive instruction to inductive, student-centered investigation.

This paradigm, pioneered by Tim Johns as Data-Driven Learning (DDL), redefines the language student as a linguistic researcher. Instead of memorizing prescriptivist rules, students use concordancers and lexical databases to identify recurring patterns, collocations, and contextual constraints in authentic texts. While the conceptual benefits of DDL are widely recognized, its practical integration into mainstream curricula—and its long-term effects on student autonomy and error self-correction—remains an active area of empirical inquiry. The issue requires research proposals in the future to investigate how corpus tools can optimize vocabulary acquisition, refine academic writing, and elevate self-directed learning competence among advanced English learners.

The theoretical foundation of this study is grounded in Data-Driven Learning (DDL) and usage-based models of language acquisition. Johns (1991) argued that the best way to internalize a language is through active discovery, summarizing the approach with the maxim: 'Every student a Sherlock Holmes.' Through the observation of concordance lines, learners engage in cognitive processing that helps them notice linguistic features that might otherwise remain overlooked in passive exposure.

Subsequent research by Boulton (2010) and Cobb (1997) has validated the efficacy of DDL for vocabulary development and grammatical refinement. Meta-analyses reveal that analyzing empirical data yields significantly higher retention rates for complex collocations than traditional rote-learning methods. However, Flowerdew (2015) notes that many corpus tools remain designed for professional researchers rather than pedagogical environments, occasionally resulting in cognitive overload for unguided students. Furthermore, while the short-term benefits of corpus use in structured environments are well-documented, longitudinal research tracing how learners autonomously apply corpus techniques to self-correct authentic drafts outside the classroom remains limited.

In addition, some restraints in the implication of DDL in English classroom should be stated; One of these is overusing adjectives while self-editing in which some learners fail to choose the right words or repeatedly use synonyms excessively. Despite of the mostly effective use of concordances, particularly low level learners still may need teacher feedback or peer editing.

The next significant factor is methodology of how DDL is implemented. There are two methods that generally adopted: Direct and indirect DDL. In direct DDL learners directly works only with corpus tools where they independently build lexico-grammar rules, sentence structures and others, while indirect method requires teacher involvement where students are presented authentic material in paper form to

© 2026 London International Monthly Conference on Multidisciplinary Research and Innovation. All Rights Reserved.

work on. Previously, some research findings proved significant advantages of the first method to gain autonomy in learning for mainly high-level learners, nevertheless, following years small scale research by Yoon and Jo (2014) concluded that self-correction rate was higher in low-level students in indirect DDL classes.

Taken together all above indicates that DDL is one of the effective and up-to-date instructional approaches that combine language learning with technology, which increases learners' autonomy by providing them the wide range of real-life texts, the correct use of words and word structures through optimizing vocabulary stock. However, there are some challenges including particularly low-level students face in DDL classrooms and lack of more empirical studies in the field. Forthcoming studies should focus on more long-term empirical studies to address the questions.

References

1. Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. Language Learning Journal, 38(1), 53–66.
2. Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? System, 25(3), 301–315.
3. Flowerdew, L. (2015). Corpus-based research and pedagogy. In Robbie, B., & Susan, H. (Eds.), The Routledge Handbook of Language Awareness (pp. 312–327). Routledge.
4. Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. English Language Research Journal, 4, 1–16.
5. O'Keeffe, A., McCarthy, M., & Carter, R. (2007). From corpus to classroom: Language use and language teaching. Cambridge University Press.
6. Sinclair, J. M. (2004). How to use corpora in language teaching. John Benjamins Publishing Company.

© 2026 London International Monthly Conference on Multidisciplinary Research and Innovation. All Rights Reserved.