pISSN:2349-5707 Volume:13,Issue 6, June -2026 eISSN:2349-5715

DATA-DRIVEN LEARNING AND THE ACQUISITION OF TECHNICAL VOCABULARY

Karimova Feruza Inoyatovna
Senior Lecturer, Department of Languages and Humanities
Andijan State Technical Institute

Abstract. The article analyzes data-driven learning (DDL) as a corpus-based approach to the acquisition of technical vocabulary in English for Specific Purposes (ESP) instruction. Conventional vocabulary teaching, organized around bilingual word lists and isolated definitions, is contrasted with the inductive examination of concordance lines, in which the learner acts as a language investigator deriving patterns of use from authentic data. Meta-analytic and review evidence is examined showing that DDL yields large effects on vocabulary and lexicogrammar and that it teaches not only meaning but collocation, colligation, and typical usage. At the same time, the research–practice gap, the cognitive demands of raw concordance data, and the limited evidence on durability of learning are discussed. Criteria for the principled classroom use of DDL in technical vocabulary instruction are formulated.

Keywords: data-driven learning, corpus linguistics, concordancing, technical vocabulary, collocation, English for Specific Purposes, lexicogrammar.

Introduction

Technical vocabulary occupies a central place in English for Specific Purposes (ESP), since command of the specialized lexicon is a precondition for participation in a professional discourse community. Conventional approaches to teaching this vocabulary rely heavily on bilingual word lists and isolated definitions, which present the learner with a decontextualized meaning but neglect the patterns in which a term actually occurs –its collocations, its grammatical environment, and its typical usage. Knowing a word, however, includes knowing the company it keeps [6; 7].

Data-driven learning, introduced by T. Johns, offers a different model [1]. Instead of receiving rules and meanings ready-made, the learner examines multiple authentic examples of a target item presented as concordance lines and induces the regularities of its use, acting, in Johns's formulation, as a language detective. The approach draws directly on the tools and findings of corpus linguistics and is closely allied to a view of language as lexicalized grammar, in which fixed and semi-fixed multiword units are basic building blocks of fluent production [6].

The aim of the present study is to analyze the evidence for the effectiveness of data-driven learning in the acquisition of technical vocabulary and to identify the conditions under which its use in the ESP classroom is methodologically justified.

Materials and Methods

The theoretical base of the study comprises Johns's original account of data-driven learning [1], the lexical approach of M. Lewis, which foregrounds collocation and the multiword unit [6], and I. S. P. Nation's framework of vocabulary knowledge, in which depth of knowledge extends well beyond form–meaning mapping [7].

The principal empirical evidence is the meta-analysis of A. Boulton and T. Cobb, which synthesized sixty-four studies representing eighty-eight independent samples [2]. This is complemented by Boulton's focused review of corpus consultation in ESP [3], by A. Chambers's analysis of the gap between corpus research and classroom practice [4], and by the survey of classroom applications of corpus analysis by Cobb and Boulton [5].

The research method combines theoretical analysis of the conceptual apparatus of data-driven learning with synthesis of meta-analytic and review evidence on its differential effects in vocabulary and lexicogrammar.

pISSN:2349-5707 Volume:13,Issue 6, June -2026 eISSN:2349-5715

Results

The meta-analysis of Boulton and Cobb established large overall effects for data-driven learning, with an effect size of d = 0.95 for control–experimental comparisons and d = 1.50 for pre-test/post-test designs [2]. The effects held across both computer-based and paper-based concordancing and were observed in particular for vocabulary and lexicogrammar, which are precisely the areas at the centre of ESP. Notably, consistent effects were found in settings where the presence of native-speaker instructors was limited, a condition characteristic of many technical institutions outside the English-speaking world.

Beyond aggregate effect sizes, the qualitative character of the learning is significant. Because concordance lines display a term in many authentic contexts at once, data-driven learning teaches not only the meaning of a technical word but its collocational and colligational behaviour and its semantic prosody –dimensions of word knowledge that word lists cannot convey [6; 7]. The inductive discovery of these patterns from authentic data is associated with deeper processing than the rote memorization of equivalents.

The reviewed work further indicates that specialized or pedagogical corpora are better suited to technical vocabulary than general-language corpora, since they expose learners to the target terms in the authentic contexts of their own discipline [3]. The corpus thus functions both as a reference resource the learner can consult and as the primary material for guided discovery.

Discussion

The strength of the evidence must be read alongside the cautions registered by the researchers themselves. Boulton and Cobb note that many of the smaller effect sizes were tied to small samples, and that the durability and transfer of learning, as measured by delayed post-testing, remain insufficiently investigated [2]. Large immediate gains are therefore not by themselves proof of lasting acquisition, and claims for the approach should be correspondingly measured.

A second consideration is the gap between research and practice. Chambers observes that, despite favourable findings, data-driven learning has not been widely adopted in ordinary classrooms, in part because teachers lack training, the tools can appear intimidating, and the preparation of materials is time-consuming [4]. Raw, unfiltered concordance data also impose a heavy cognitive load, especially on lower-level learners. The practical answer is a mediated, scaffolded form of the approach in which the teacher prepares concordance materials suited to the learners' level rather than abandoning them to the unprocessed output of a corpus tool.

For technical education in particular, the fit is nonetheless strong. Where qualified native-speaker instructors are scarce, a well-chosen specialized corpus supplies authentic disciplinary input independent of the teacher's own intuitions, and it does so for exactly the lexicogrammatical patterning that distinguishes competent professional writing from a merely correct rendering of dictionary equivalents. The role of the teacher shifts accordingly, from supplier of definitions to designer of guided discovery.

Conclusion

Data-driven learning is an evidence-backed approach to technical vocabulary, supported by large effect sizes and well suited to settings in which authentic input cannot be supplied by native-speaker instructors. Its classroom value, however, depends not on the mere use of a corpus but on the design of the activity. The teacher introducing data-driven learning in an ESP course should test the approach against three criteria. Does the corpus material expose learners to authentic disciplinary usage –collocation, colligation, typical patterning –rather than to isolated meaning alone? Is the concordance data scaffolded to the learners' level rather than overwhelming them with unfiltered output? Is the teacher equipped to mediate the process of discovery rather than leaving learners alone with raw data? When these conditions are met, data-driven learning becomes a methodologically justified means of building technical vocabulary in the ESP classroom.

pISSN:2349-5707 Volume:13,Issue 6, June -2026 eISSN:2349-5715

References

1. Johns T. Should You Be Persuaded: Two Samples of Data-Driven Learning Materials // English Language Research Journal. – 1991. – Vol. 4. – P. 1–16.
2. Boulton A., Cobb T. Corpus Use in Language Learning: A Meta-Analysis // Language Learning. – 2017. – Vol. 67. – No. 2. – P. 348–393.
3. Boulton A. Corpus Consultation for ESP: A Review of Empirical Research // Corpus-Informed Research and Learning in ESP: Issues and Applications / Eds. A. Boulton, S. Carter-Thomas, E. Rowley-Jolivet. – Amsterdam: John Benjamins, 2012. – P. 261–291.
4. Chambers A. Towards the Corpus Revolution? Bridging the Research–Practice Gap // Language Teaching. – 2019. – Vol. 52. – No. 4. – P. 460–475.
5. Cobb T., Boulton A. Classroom Applications of Corpus Analysis // The Cambridge Handbook of English Corpus Linguistics / Eds. D. Biber, R. Reppen. – Cambridge: Cambridge University Press, 2015. – P. 478–497.
6. Lewis M. The Lexical Approach: The State of ELT and a Way Forward. – Hove: Language Teaching Publications, 1993.
7. Nation I. S. P. Learning Vocabulary in Another Language. – 2nd ed. – Cambridge: Cambridge University Press, 2013.