

DATA-DRIVEN LEARNING AND ACADEMIC REGISTER AWARENESS: A CORPUS-INTEGRATED APPROACH FOR FUTURE ENGLISH TEACHERS

Karimov Asilbek Rayimjon o'g'li

Student of the Department of Theory and Practice of the English Language Chirchik State Pedagogical University

Golovko Yaroslav Vladimirovich

Scientific adviser: Senior teacher Department of Theory and Practice of the English Language Chirchik State Pedagogical University

Abstract: *This article investigates the development of academic register awareness and lexical precision through data-driven learning (DDL) methodology in the training of future English teachers. Drawing on experimental research conducted with B2-level EFL students at Chirchiq State Pedagogical University using the COCA concordancer, the study demonstrates that register awareness and lexical precision are closely interconnected dimensions of academic writing competence that benefit substantially from corpus-integrated instruction. Students who engaged with concordance data showed improved ability to distinguish between conversational and academic lexical alternatives, suggesting that corpus tools facilitate the development of stylistic sensitivity alongside lexical accuracy. The findings support the argument that corpus-integrated instruction should be systematically embedded in teacher education curricula.*

Keywords: *register awareness, lexical precision, data-driven learning, academic writing, corpus tools, future English teachers.*

Annotatsiya: *Ushbu tadqiqotda ma'lumotlarga asoslangan o'rganish (DDL) metodologiyasi yordamida bo'lajak ingliz tili o'qituvchilarida akademik registr va leksik aniqlikni rivojlantirish masalasi ko'rib chiqiladi. COCA konkordansidan foydalangan holda o'tkazilgan eksperimental tadqiqot natijalariga ko'ra, register haqida xabardorlik va leksik aniqlik o'rtasidagi bog'liqlik korpusga asoslangan ta'limning eng samarali natijalaridan biri hisoblanadi. Talabalar so'zlashuvga xos va akademik uslubiy alternativlarni taqqoslash orqali rasmiy akademik iboralarni yanada aniq farqlashni o'rgandilar. Tadqiqot natijalari kelajak pedagoglar tayyorlash dasturlariga korpus vositalarini integratsiya qilishning zarurligini tasdiqlaydi.*

Kalit so'zlar: *register haqida xabardorlik, leksik aniqlik, ma'lumotlarga asoslangan o'rganish, akademik yozuv, korpus vositalari, bo'lajak o'qituvchilar.*

Introduction: Academic register represents one of the most complex dimensions of writing competence for EFL learners transitioning from general English instruction to academic writing courses. The distinction between conversational and academic language involves not only vocabulary selection but also syntactic choices, discourse organization, and adherence to disciplinary norms. Many learners, even those with relatively high general proficiency, struggle to produce writing that consistently reflects the formality, objectivity, and precision characteristic of academic discourse.

Lexical precision, closely related to register awareness, refers to the ability to select vocabulary that accurately and appropriately expresses the intended meaning within a given academic context. Unlike general communicative competence, lexical precision in academic writing requires familiarity with the semantic nuances of near-synonymous terms, an understanding of their collocational behavior, and sensitivity to their stylistic connotations. Traditional instruction often fails to develop this dimension of competence, as coursebook exercises rarely expose learners to the full range of contextual variation that characterizes authentic academic language use.

Corpus tools, particularly concordancers such as COCA, offer a promising solution by enabling learners to directly observe how lexical items function in real academic texts. By comparing informal and formal lexical alternatives within authentic contexts, learners can develop both register awareness and lexical precision simultaneously. The present study reports on an experimental intervention designed to evaluate the effectiveness of this approach among future English teachers.

Methods

The experimental study was carried out with 32 third-year English Language and Literature students at Chirchiq State Pedagogical University, all classified as B2-level EFL learners. The intervention consisted of an 80-minute corpus-integrated lesson using a structured worksheet that guided students through a series of concordance-based tasks targeting academic register and lexical precision. Tasks included comparing informal and academic lexical alternatives such as a lot of versus a significant number

of, get better versus improve, and bad sources versus unreliable sources, as well as identifying register-appropriate vocabulary in concordance lines and applying corpus evidence to revise sentences for academic style.

Students completed a 20-item reflective questionnaire immediately after the lesson, rating their perceived improvement in each targeted subskill on a 1-10 scale. Six experienced EFL academic writing teachers also evaluated the worksheet using a 23-item questionnaire and five open-ended questions. The data were analyzed descriptively, with particular attention to the register and lexical precision items in both questionnaires.

Results and Discussion

Both register awareness and lexical precision received mean scores of 7.78 in the student questionnaire, placing them in the middle of the effectiveness ranking but still reflecting strongly positive perceptions. Qualitative analysis of student reflections revealed that the corpus-based comparison of informal and academic alternatives was particularly valuable, as it allowed learners to see not merely that certain words are more formal, but to observe precisely how and where they are used by proficient academic writers.

Teacher evaluations confirmed the interconnected nature of these two subskills. Teachers emphasized that corpus-based instruction helps learners move beyond intuitive stylistic judgments toward evidence-based lexical choices. Open-ended teacher responses highlighted that the ability to observe high-frequency academic vocabulary in authentic concordance lines enables students to internalize register norms more reliably than prescriptive stylistic rules. Several teachers noted that the worksheet was particularly effective in addressing the common tendency among EFL learners to rely on vague or conversational language when writing academically.

The finding that register awareness and lexical precision scored equally in student perceptions is pedagogically significant, as it suggests that learners themselves recognize the close relationship between these dimensions. Corpus tools appear to bridge the gap between knowing that a word is formal and understanding when and how it is appropriately used, which is the critical distinction between passive stylistic awareness and active academic language competence.

Conclusion

This study demonstrates that corpus-integrated instruction through data-driven learning activities is effective in developing both register

awareness and lexical precision among future English teachers. By engaging with authentic COCA concordance data, B2-level EFL students developed greater sensitivity to the stylistic and semantic differences between conversational and academic language, which are essential competencies for producing high-quality academic writing. The pedagogical implications are clear: corpus tools should be systematically integrated into academic writing instruction in teacher education programs, with activities specifically designed to address register awareness and lexical precision alongside other dimensions of academic discourse competence. Future research should examine the long-term effects of corpus-based instruction on independent writing performance in authentic academic assessment contexts.

REFERENCES:

1. Hyland K. Academic Discourse: English in a Global Context. – London: Continuum, 2009.
2. Coxhead A. A new academic word list // TESOL Quarterly. – 2000. – Vol. 34(2). – P. 213–238.
3. Biber D., Gray B. Grammatical complexity in academic English. – Cambridge: Cambridge University Press, 2016.
4. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012.
5. Hunston S. Corpora in Applied Linguistics. – Cambridge: Cambridge University Press, 2002.
6. Johns T. Data-driven learning: The perpetual challenge // Computer Assisted Language Learning. – 1991. – Vol. 4, № 2–3. – P. 107–117.
7. Granger S. Learner English on Computer. – London: Longman, 1998.