



**Nilufar BARNOYEVA**

Associate professor (PhD), Department of Foreign Languages Tashkent University of Information Technologies  
E-mail: barnoyevanilufar1984@gmail.com

Based on the review by D.A.Radjabova, Associate Professor of TUIT named after Muhammad al Khwarizmi, PhD

### ENHANCE ICT STUDENTS' LINGUISTIC COMPETENCE THROUGH CORPUS LINGUISTICS

Annotation

This article explores the potential of Corpus Linguistics as an effective tool to enhance the linguistic competence of students in Information and Communication Technology (ICT) fields. Given the increasing demand for proficient English usage in technical and professional contexts, this study emphasizes the integration of corpus-based methodologies in language learning. The research outlines theoretical foundations, practical applications, and evaluates the outcomes of implementing corpus-based instruction for ICT students.

**Key words:** Corpus Linguistics, ICT students, linguistic competence, Data-Driven Learning.

### ПОВЫШЕНИЕ ЯЗЫКОВОЙ КОМПЕТЕНЦИИ СТУДЕНТОВ ИКТ СРЕДСТВАМИ КОРПУСНОЙ ЛИНГВИСТИКИ

Аннотация

В данной статье рассматривается потенциал корпусной лингвистики как эффективного инструмента для повышения языковой компетенции студентов в области информационно-коммуникационных технологий (ИКТ). Учитывая возрастающий спрос на владение английским языком в технических и профессиональных контекстах, исследование акцентирует внимание на интеграции корпусных методик в процесс изучения языка. В работе раскрываются теоретические основы, практические применения и проводится оценка результатов внедрения корпусно-ориентированного обучения для студентов ИКТ.

**Ключевые слова:** Корпусная лингвистика, студенты ИКТ, языковая компетенция, обучение на основе данных.

### TIL KORPUSI ASOSIDA AKT TALABALARINING LINGVISTIK KOMPETENSIYASINI RIVOJLANTIRISH

Annotatsiya

Ushbu maqolada til korpusi Axborot va Kommunikatsiya Texnologiyalari (AKT) yo'nalishida tahsil olayotgan talabalarning lingvistik kompetensiyasini rivojlantirishda samarali vosita sifatida ko'rib chiqiladi. Texnik va sohaga oid kontekstlarda ingliz tilidan samarali foydalanishga bo'lgan talabning ortib borayotganligini inobatga olib, tadqiqotda til o'rganish jarayoniga korpusga asoslangan metodologiyalarni tadbiiq qilishning ahamiyati ta'kidlanadi. Maqolada nazariy asoslar, amaliy qo'llanmalar yoritilib, AKT talabalari uchun korpusga asoslangan ta'limni joriy etish natijalari baholanadi.

**Kalit so'zlar:** Til korpusi, AKT talabalari, lingvistik kompetensiya, Data-driven learning (ma'lumotlarga asoslangan ta'lim).

In the modern digital world, ICT specialists are expected to have strong linguistic skills to communicate effectively in global contexts. Linguistic competence, in this context, encompasses the ability to use accurate and context-appropriate vocabulary, understand syntactic structures, interpret professional discourse, and apply pragmatic conventions in both written and spoken communication. This includes reading and writing technical documentation, understanding error messages, commenting on code effectively, collaborating in multicultural teams, and delivering presentations or instructions in English.

English, being the lingua franca of science and technology, necessitates a high level of proficiency, especially in technical documentation, programming commentary, collaborative projects, and international communication. Traditional methods of language teaching often fall short in preparing students for these real-world requirements, as they frequently rely on scripted dialogues and decontextualized grammar exercises.

Corpus Linguistics (CL), as an empirical approach to language study, provides access to authentic language use and allows learners to investigate real-life usage patterns. Through exposure to large collections of actual texts—ranging from technical manuals and helpdesk conversations to user guides,

source code comments, and IT forum discussions—learners are immersed in naturally occurring language relevant to their field. This allows them to observe how language is structured and used in meaningful contexts.

One of the core strengths of CL lies in its ability to promote inductive learning through empirical observation. Learners are encouraged to formulate hypotheses, test them against real-world data, and discover linguistic norms on their own. For instance, through the analysis of concordance lines, students might identify that the phrase “open source” frequently appears with verbs like “use,” “develop,” and “support,” or that conditional structures are commonly used in technical instructions and troubleshooting texts.

CL also enhances understanding of lexical bundles, register variation, and pragmatic usage. For ICT students, who must navigate different communication genres—such as bug reports, documentation, technical proposals, and email correspondence—corpora expose them to stylistic differences and allow them to tailor their language use accordingly.

Furthermore, corpus tools support metalinguistic awareness and self-correction. By examining learner corpora or comparing learner output with expert models, students can detect gaps in their usage, such as missing prepositions, awkward collocations (e.g., “use the internet” instead of “surf

the internet” or “make a backup” instead of the more accurate “create a backup”), or incorrect verb tenses in procedural writing. Additional ICT-specific collocations that students explored included “configure settings,” “establish connection,” “retrieve data,” “encrypt files,” “monitor performance,” “execute command,” “allocate memory,” and “integrate systems.” Recognizing these patterns helped learners internalize conventional usage and apply them appropriately in technical writing and spoken communication.

Corpus-based instruction thus supports all aspects of linguistic competence—lexical, grammatical—by reinforcing exposure to authentic usage, encouraging discovery-based learning, and fostering critical reflection on language. This paper aims to analyze the implementation of CL as a means to enhance linguistic competence among ICT students by aligning language instruction with the authentic demands of their future professional contexts.

The study employed a mixed-methods research design combining quantitative and qualitative approaches. A pedagogical experiment was conducted over a six-week instructional period to assess the impact of corpus-based instruction on ICT students’ linguistic competence. Pre-test and post-test measures were used to evaluate students’ progress, while classroom observations and task analyses provided qualitative insights.

The participants were undergraduate first-year ICT students enrolled in an English language course at a higher education institution. The students had an intermediate level of general English proficiency but limited experience with professional ICT discourse. All participants received the same number of instructional hours and followed the same syllabus, with corpus-based activities integrated into the experimental instruction.

The instructional materials were based on authentic corpus data drawn from large general and specialized corpora, such as the News on the Web (NOW) and collections of ICT-related texts. Concordance and frequency analysis tools were used to identify typical collocations and usage patterns of key ICT terms (e.g., system, network, data, software).

Examples of corpus use:

Using the NOW corpus, students generated concordance lines for the noun system and observed recurring patterns such as operating system, upgrade the system, system failure, and secure system. From these examples, students inferred that system frequently co-occurs with evaluative adjectives (secure, reliable) and process-oriented verbs (upgrade, maintain, deploy).

Frequency lists and collocation statistics (e.g., MI-score, raw frequency) were applied to the noun data, allowing students to identify typical combinations such as collect data, process data, data breach, and data storage. Students compared these findings with literal translations from their first language and discussed mismatches.

Concordance (e.g., built-in corpus interfaces) and frequency analysis tools were used to visualize patterns of use. Students learned to read concordance lines, interpret left and right contexts, and distinguish between general and domain-specific meanings.

The teaching methodology followed the principles of Data-Driven Learning (DDL), in which students actively explored corpus data rather than receiving ready-made explanations.

Examples of DDL tasks:

Concordance analysis (receptive): Students examined 20–30 concordance lines for the word network and answered guided questions such as: “Which verbs most frequently

precede network?” and “What adjectives describe network performance?”

Collocation matching (receptive): Based on corpus findings, students matched verbs (deploy, configure, monitor) with nouns (software, system, network) and justified their choices using corpus evidence.

Pattern identification (guided discovery): Students compared patterns like install software vs. set software and concluded why only one option is acceptable in professional ICT discourse.

Productive tasks: Learners used identified collocations to write short technical descriptions (e.g., explaining how to secure a network or process data in an application) and to complete gap-fill tasks derived from authentic corpus sentences.

Both receptive (analysis and interpretation of corpus data) and productive (sentence construction and short professional texts) exercises were systematically integrated, ensuring a gradual transition from observation to independent language use.

The teaching methodology followed the principles of Data-Driven Learning (DDL), in which students actively explored corpus data rather than receiving ready-made explanations. Tasks included concordance analysis, collocation matching, pattern identification, and guided discovery activities. Both receptive (analysis and interpretation) and productive (sentence construction and short texts) exercises were implemented.

Data-driven learning (DDL) activities where students analyzed concordance lines, word frequency, and collocations relevant to ICT terminology. These activities helped students explore the most frequently used technical terms, observe how these terms function in various syntactic structures, and identify patterns of usage in professional contexts. For instance, students learned to distinguish between different usages of “run,” “compile,” and “deploy” in software documentation and forum discussions. The DDL tasks were structured in stages: initial guided exploration, followed by independent analysis and collaborative group discussions to share insights. Other examples included analyzing collocations such as “install software,” “write code,” “debug application,” “access database,” “launch program,” “set parameters,” and “update drivers.” By identifying these common word pairings and their usage in authentic contexts, students gained a clearer understanding of domain-specific lexical bundles and phraseology. The DDL tasks were structured in stages: initial guided exploration, followed by independent analysis and collaborative group discussions to share insights.

The DDL tasks were structured in stages to gradually develop student autonomy and deepen their understanding of language use in ICT contexts. In the initial guided exploration stage, the instructor demonstrated how to use corpus tools such as AntConc or Sketch Engine to search for specific ICT-related terms and analyze their usage in context. Students were shown how to interpret concordance lines, recognize patterns of word usage, and differentiate between literal and idiomatic expressions.

The results of the study indicate a positive impact of corpus-based instruction on ICT students’ linguistic competence. Quantitative analysis of test scores revealed a noticeable improvement in students’ ability to recognize and correctly use ICT-related collocations after the instructional intervention.

Students demonstrated increased accuracy in combining technical nouns with appropriate verbs and adjectives (e.g., deploy software, maintain a system, secure data). In addition, qualitative observations showed that students

became more confident in interpreting unfamiliar lexical items by analyzing their context of use in corpus data.

Another significant outcome was the development of students' analytical and critical thinking skills. While working with concordance lines, students learned to compare examples, identify recurring patterns, and justify their conclusions with linguistic evidence. This shift from passive reception to active discovery contributed to higher learner engagement and autonomy.

The findings of the study confirm that Corpus Linguistics can serve as an effective methodological framework for enhancing ICT students' linguistic competence. Exposure to authentic language data enables learners to overcome limitations of traditional materials and to develop a more realistic understanding of professional English usage.

The observed improvement in collocational competence supports previous research emphasizing the importance of corpus-based approaches in ESP instruction. Moreover, the integration of DDL tasks aligns with constructivist learning principles, as students actively construct knowledge through exploration and analysis. For ICT students, who are generally accustomed to analytical and problem-solving tasks, corpus-based activities appear particularly suitable.

However, the study also revealed certain challenges, such as the initial difficulty students faced when interpreting concordance data without sufficient guidance. This suggests that corpus-based instruction should be carefully scaffolded and adapted to learners' proficiency levels.

In the next stage, independent analysis, students selected their own set of terms or collocations from the corpus

and performed individual searches. They documented findings regarding frequency, collocation strength, and grammatical patterns, building their personal vocabulary and usage banks. This stage encouraged critical thinking and allowed students to take ownership of their learning.

Finally, in the collaborative group discussions, students presented their findings, compared results, and discussed variations in usage across different text types such as manuals, technical blogs, or forum posts. This peer interaction helped reinforce their discoveries and introduced them to additional usage patterns that may have been missed individually, fostering a deeper collective understanding of technical language use.

The integration of Corpus Linguistics into language instruction for ICT students demonstrates significant language learning value. By grounding language learning in authentic, data-driven analysis, students acquire not only technical vocabulary and grammatical precision but also critical language awareness and adaptive communication skills suited for global professional environments. Corpus-based methods offer an empirical, learner-centered approach that enhances all dimensions of linguistic competence—lexical, syntactic, pragmatic, and discourse-related. This research supports the broader adoption of corpus tools in ESP (English for Specific Purposes) curricula and encourages educators to incorporate DDL strategies that align language instruction with real-world technical demands. Future research may explore long-term impacts, cross-disciplinary implementations, and the development of specialized corpora tailored to emerging ICT subfields.

#### REFERENCES

1. Белов, С. А. Корпусная лингвистика: Учебное пособие. — М.: Логос, 2020. — 256 с.
2. Flowerdew, L. Corpora and Language Education. — New York: Palgrave Macmillan, 2012. — 202 p.
3. Hunston, S. Corpora in Applied Linguistics. — Cambridge: Cambridge University Press, 2002. — 189 p.
4. Малышева, И. А. Использование корпусов текста в обучении профессиональному английскому языку. // Язык и культура. — 2021. — №3. — С. 55–62.
5. Sinclair, J. Corpus, Concordance, Collocation. — Oxford: Oxford University Press, 1991. — 160 p.
6. Tribble, C., & Jones, G. Concordances in the Classroom: A Resource Book for Teachers. — London: Longman, 1997. — 110 p.
7. Ушакова, Т. Н. Современные подходы к обучению языку профессионального общения. — СПб.: Изд-во СПбГУ, 2019. — 214 с.
8. Canale, M., & Swain, M. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. // Applied Linguistics. — 1980. — Vol. 1, No. 1. — P. 1–47.
9. Bachman, L. F. Fundamental Considerations in Language Testing. — Oxford: Oxford University Press, 1990. — 408 p.
10. Celce-Murcia, M. (Ed.). Teaching English as a Second or Foreign Language. — Boston: Heinle & Heinle, 2001. — 446 p.