ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-515-x, Impact Factor: 9.23

American Academic publishers, volume 6, issue 01,2026

Journal: https://www.academicpublishers.org/journals/index.php/ijai

THE IMPACT OF CORPUS-BASED LINGUISTIC ANALYSIS ON PROFESSIONAL LANGUAGE COMPETENCE

Gulkhumor Kakhkhorova Sulaymonjon kizi
Independent researcher, Department of Uzbek Language and Language Teaching, Fergana State Technical University, Fergana, 150102, Uzbekistan
E-mail: gulhumor91.uz@gmail.com

Abstract. This study examines the effectiveness of corpus-based linguistic analysis in enhancing the professional language competence of 60 undergraduate artificial intelligence students at Fergana State Technical University. Over the course of a semester, students in the experimental group engaged in hands-on analysis of authentic AI-related texts using specialized corpus linguistics tools, namely AntConc and Sketch Engine. These tools enabled students to explore language data through frequency lists, lemmatization, collocation identification, and topic modeling, providing deep insights into the usage of domain-specific vocabulary and phraseology. In contrast, the control group followed the traditional curriculum, which focused on theoretical instruction and instructor-led exercises without exposure to corpus technologies. Pre- and post-intervention assessments measured students' proficiency in using technical terminology, constructing professional sentences, and recognizing collocations relevant to the AI field.

Keywords: corpus analysis, professional language, AI education, topic modeling, experimental study

Introduction

In today's globalized and technology-driven workplace, the ability to communicate effectively in professional and technical language is a core competency for artificial intelligence (AI) specialists. Undergraduate AI students must not only master the technical aspects of their field but also acquire the language skills necessary to articulate complex ideas, collaborate with international teams, and document their work according to industry standards. However, traditional language instruction in higher education often lacks the authenticity and specificity required to meet these professional needs. Typical curricula are limited to textbook examples and teacher-centered methods, which may not adequately expose students to the real-world discourse and terminology prevalent in AI research and industry. As a result, graduates can face difficulties in conveying technical information precisely and fluently, potentially hindering their professional growth.

To address these issues, language educators and researchers have turned to corpus linguistics—a branch of linguistics that relies on large, computerized collections of authentic texts (corpora) to analyze actual language use. Corpus-based approaches enable direct engagement with real, field-specific language data, allowing students to observe patterns, frequencies, and contextual usage of terminology and structures that textbooks often overlook. This research investigates whether incorporating corpus-based linguistic analysis into the language curriculum can more effectively develop the professional language competence of AI students compared to conventional methods. Specifically, it examines how exposure to authentic AI corpora and hands-on data analysis with tools like AntConc and Sketch Engine influence students'

ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

acquisition of domain-specific vocabulary, phraseology, and written communication skills. The study aims to provide evidence-based recommendations for modernizing language education in technical disciplines and preparing graduates for the demands of the AI industry.

Methods

The study was conducted with 60 undergraduate students enrolled in the artificial intelligence program at Fergana State Technical University. Participants were randomly assigned to either an experimental group (n=30) or a control group (n=30) to ensure balanced representation and minimize bias. The experimental group took part in weekly workshops where they were introduced to the fundamentals of corpus linguistics and trained to use AntConc and Sketch Engine—two widely recognized tools for corpus analysis. Throughout the semester, students worked with a curated corpus of AI research articles, technical documentation, and industry reports. They conducted frequency analyses to identify the most common terms and phrases within the AI domain, used lemmatization to group word variations, explored collocations to uncover frequently co-occurring lexical items, and performed topic modeling to detect underlying themes and trends in professional language.

Assignments required students to apply these findings to their own writing tasks, such as technical summaries, project proposals, and documentation, with an emphasis on using authentic language patterns discovered in the corpora. The control group, in contrast, followed a standard curriculum that included lectures, textbook exercises, and teacher-led discussions on professional English, without any exposure to corpus tools or real-world data. Both groups completed pre-tests and post-tests that measured their knowledge and accurate use of technical vocabulary, ability to construct professional sentences, and recognition of domain-specific collocations. Additionally, the experimental group provided qualitative feedback through surveys and short interviews regarding their experiences, perceived challenges, and learning outcomes with corpus-based instruction.

Results

The findings revealed a significant difference in the professional language competence gained by the experimental group compared to the control group. After one semester, the experimental group demonstrated a 30% increase in the correct and contextually appropriate use of AI-related terminology and collocations, as measured by standardized post-test results. Their written assignments reflected a broader and more precise use of technical vocabulary, as well as greater syntactic complexity—often mirroring the authentic language patterns observed in the analyzed corpora. For example, students became adept at incorporating multi-word expressions, technical jargon, and standard phrases that are common in AI literature but rarely found in traditional textbooks.

In contrast, the control group's improvement was more modest, with only a 12% rise in the same metrics. Their language use generally remained within the scope of textbook examples and lacked the richness and authenticity observed in the experimental group's work. Qualitative feedback from the experimental group further emphasized the pedagogical value of corpus analysis; students reported increased motivation, deeper engagement, and a sense of empowerment from actively discovering and applying real professional language. Many noted that analyzing authentic texts made abstract linguistic concepts more tangible and provided concrete models for their own writing. Some students initially faced technical challenges in using the software, but most adapted quickly and expressed a desire to continue using corpus tools in future studies. Overall, the integration of corpus-based analysis not only enhanced

ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

students' linguistic competence but also fostered more independent, research-oriented learning habits.

Discussion

The outcomes of this experimental study strongly support the integration of corpus-based linguistic analysis into language education for artificial intelligence students. By working directly with authentic, field-specific corpora, students not only acquire the vocabulary and phraseology essential for professional communication but also develop a more nuanced understanding of the conventions and expectations of their discipline. The substantial improvement in the experimental group's use of technical terms, collocations, and syntactic structures demonstrates that corpus-driven instruction bridges the gap between academic learning and professional practice far more effectively than traditional textbook-based methods.

Furthermore, the hands-on nature of corpus analysis encourages students to become active investigators of language, fostering critical thinking and self-directed learning skills that are highly valued in both academia and industry. The positive qualitative feedback highlights the motivational and confidence-building effects of engaging with real data, as students witness their own progress and gain practical insights into language variation and use. However, the study also acknowledges certain limitations, such as the initial learning curve associated with mastering corpus tools and the need for instructor guidance in interpreting complex linguistic patterns.

Future research and practice should focus on refining corpus-based instructional models, providing robust technical support, and integrating corpus analysis activities across a wider range of professional and technical disciplines. Overall, this study demonstrates that corpus-based linguistic analysis is a highly effective strategy for enhancing the professional language competence of AI students, equipping them with the communication skills necessary for successful participation in the global technology sector.

.

Reference.

1. Qahhorova, G. (2023). Ingliz tilida tinglash mahoratini yaxshilash usullari. *Journal of Science-Innovative Research in Uzbekistan, 1*(9), 1249-1254.
2. Qahhorova, G. (2023). O'quvchilarning tinglash malakasini oshirish uchun diktantlardan samarali foydalanish usullari. *Journal of Science-Innovative Research in Uzbekistan, 1*(9), 1255-1260.
3. Gulkhumor Kakhkhorova. (2025). Verbalizing code and model architecture and a linguistic-analytical technique for AI students. *Shokh Articles Library, 1*(1).
4. Gulkhumor Kakhkhorova. (2025). The impact of a linguistic-analytical approach on AI students' professional speech skills. *Shokh Articles Library, 1*(1).