

Talabalar yozma kompetensiyasini rivojlantirishda DDL korpus yondashuvining muhim xususiyati sifatida autentiklik

Radjabova Gulnoza Giyosiddinovna

rad.gulnoza@gmail.com

PhD, v.b. dotsent

O'zbekiston Davlat Jahon Tillari Universiteti

Annotatsiya. Ushbu maqolada zamonaviy amerika ingliz tili korpusi (COCA) DDL korpus yondashuvi asosida talabalar yozma ko'nikmalarini rivojlantirishda samarali vosita sifatida ko'rib chiqiladi. COCA korpusining autentiklik, ochiqlik, vakillik va funktsionallik kabi afzalliklari mustaqil o'rganish va tilni to'g'ri qo'llashda muhim ahamiyatga ega. Autentik materiallar o'quvchilarning motivatsiyasini, tanqidiy fikrlashini va til me'yorlarini tushunishini oshirishga yordam beradi. COCA korpusidan akademik yozuvda foydalanish ingliz tilidagi matnlarni grammatik va uslubiy jihatdan to'g'ri tuzish ko'nikmasini yuksaltiradi.

Kalit so'zlar: COCA, akademik yozuv, korpus lingvistikasi, DDL yondashuvi, autentik materiallar, til o'rganish

Аутентичность как значимая характеристика корпусного подхода DDL для совершенствования письменной компетенции студентов

Раджабова Гулноза Гиёсиддиновна

rad.gulnoza@gmail.com

PhD, доцент

Узбекский государственный университет мировых языков

Аннотация. Данная статья посвящена использованию корпуса современного американского английского языка (COCA) как эффективного инструмента для совершенствования академического письма студентов в рамках корпусного подхода DDL. Рассматриваются преимущества COCA, такие как аутентичность, доступность, репрезентативность и функциональность, способствующие самостоятельному обучению и точному употреблению языка. Подчеркивается важность аутентичных материалов для повышения мотивации обучающихся, развития критического мышления и понимания языковых норм. Интеграция COCA в процесс обучения академическому письму способствует повышению грамотности и стилиевой точности написания текстов на английском языке.

Ключевые слова: COCA, академическое письмо, корпусная лингвистика, подход DDL, аутентичные материалы, изучение языка

Authenticity as a Significant Feature of the Corpus-Based DDL Approach for Improving Students' Writing Competence

Radjabova Gulnoza Giyosiddinovna

rad.gulnoza@gmail.com

PhD, Associate Professor

Uzbekistan State World Languages University

Annotation. This article explores the use of the Corpus of Contemporary American English (COCA) as an effective tool for enhancing students' academic writing skills through the Data-Driven Learning (DDL) approach. It highlights the benefits of COCA, including its authenticity, accessibility, representativeness, and functionality, which support independent learning and accurate

language use. The study emphasizes the value of authentic language materials in improving learners' motivation, critical thinking, and understanding of linguistic norms. The integration of COCA in academic writing instruction contributes to higher proficiency in constructing grammatically and stylistically appropriate texts in English.

Keywords: COCA, academic writing, corpus linguistics, DDL approach, authentic materials, language learning

Introduction

At the current stage of linguistic development, corpus linguistics is most often regarded as an “effective tool for scientific research,” offering open access to materials from various linguistic fields such as lexicology, grammar, phonetics, translation theory, stylistics, and so on. However, it is important to note that the creation and development of corpora, as one of the most crucial tools in corpus linguistics, progress unevenly across different languages. For some languages, such as English, German, Chinese, or Japanese, massive and representatively annotated corpora already exist or are actively being developed. Meanwhile, for other languages, including Uzbek, the creation of representative and authentic corpora is still at an early stage. The situation is even more challenging for multilingual (parallel or comparative) corpora, which prevents the full realization of “the enormous potential of corpora and corpus technologies in the process of various linguistic studies” (Nagel, 2008; 58).

It should be noted, however, that over the past decades, extensive text corpora have been used by researchers as materials reflecting the realities of language in its natural state and use. These have undoubtedly had a significant impact on enhancing the quality of language resources, such as dictionaries (Longman, Oxford, Collins) and English grammar textbooks (Longman Grammar of Spoken and Written English). In recent decades, with the advancement of computer technologies, numerous new language learning approaches have emerged, one of which is Computer-Assisted Language Learning (CALL). Although CALL is gaining increasing popularity in language education, one of its specific methods – the corpus-based Data-Driven Learning (DDL) approach still remains “in the shadows” and is not widely used among foreign language teachers.

Research methods

This research focuses on the issue of how using the COCA corpus as a source of authentic materials can best support academic writing instruction in a foreign language in order to improve the writing competence of students at language universities. Specifically, we concentrated on the relationship between the types of students' writing errors and how easily these can be corrected using data available in COCA.

The **Corpus of Contemporary American English (COCA)** allows for the study of linguistic realities and is user-friendly and accessible. While there are other free corpora, COCA available online since 2008 is the largest free corpus of the English language and offers significant advantages over other corpora in studying grammar, vocabulary, and improving writing competence (Radjabova G.G., 2024). The table below (Table 1.) presents the main features of COCA, which is a mixed corpus that includes both written and spoken texts, available in audio, video, and with transcriptions. COCA is multilingual, as data is presented in English with translation options into other languages. The corpus covers a wide range of genres and registers, including conversational, journalistic, academic, literary texts, dialects, terminology, and online communication. Its data is regularly updated with current texts (more than 25 million words annually), which allows it to reflect the contemporary state of the English language. COCA supports various types of searches, including morphological, syntactic, semantic, and prosodic, and is designed to monitor current language use.

Table 1.

Typological Characteristics of the COCA Corpus

Typological Feature	COCA Characteristics
Types of Linguistic Data	Mixed – includes written and spoken texts; spoken texts are in video/audio with hyperlinks to source materials and attached transcriptions.
Language	Multilingual – all data is in English, but selected words and phrases can be translated into other languages.
Registers/Genres	Conversational, journalistic, academic, literary, dialectal, terminological, online/blog communication;
Updates	Dynamic – “the corpus contains over one billion words and is updated annually (25+ million words added each year)”
Search Capabilities	Morphological, syntactic, semantic, prosodic;
Purpose	Monitoring – this corpus was developed to “reflect current English language usage” (Radjabova G.G., 2023) and includes both written and transcribed spoken materials;

As shown in the table, this corpus is extensive, dynamic, authentic, and multi-genre, which aligns with the goals of our research. Due to its enormous size and the wide variety of genres of “written material from websites, newspapers, magazines, and books published worldwide, as well as spoken material from radio, television, and everyday conversations” (Davies, 2009; 25-37) that it covers, and the fact that new data is constantly added to monitor language changes reflecting societal and global developments, COCA is an excellent source of authentic material for a wide range of studies. It is of particular interest to our research as a source of authentic materials for improving students’ writing competence because:

First, the large volume of COCA provides sufficient insight into English vocabulary and grammar, allowing for an accurate understanding of word frequency and usage in real-life contexts. Second, working with COCA is so convenient that users do not need any special linguistic knowledge or computer skills. **Thirdly**, the advantage of COCA is that it is a balanced corpus in terms of representativeness, as it reflects various registers and genres in equal proportions, which allows for a more objective understanding of language usage in different contexts. **Fourthly**, the data in COCA are tagged and annotated, which makes it easier to search for specific linguistic phenomena, such as collocations, word frequency, and grammatical structures, all of which are essential for improving writing skills. **Fifthly**, COCA enables users to analyze language use over time, providing diachronic data that reflect changes and trends in language development. This allows students not only to learn current norms of written English but also to understand the dynamics of its evolution.

Results and discussions

Thus, due to its authenticity, accessibility, representativeness, and functionality, COCA is one of the best tools for applying the DDL (Data-Driven Learning) approach in foreign language instruction. Authenticity, in this context, is defined as the degree to which the materials used in the learning process reflect real language usage. According to many researchers, the use of authentic materials in language learning significantly increases students’ motivation, allows them to better understand the target language, and fosters the development of critical thinking and analytical skills. In the context of teaching academic writing, the authenticity of corpus materials ensures that students work with real examples of written language, which helps them to improve their competence in constructing grammatically correct and stylistically appropriate texts.

In this regard, our research emphasizes that the integration of COCA into the academic writing teaching process not only helps to correct students' errors in writing but also promotes the development of independent learning skills, since working with the corpus requires students to analyze data, make conclusions, and apply the acquired knowledge in practice. Therefore, the implementation of the DDL approach based on authentic corpus materials, such as those found in COCA, is a promising direction for improving the quality of foreign language instruction in higher education institutions, especially in language universities, where the goal is to train highly competent specialists capable of effectively using a foreign language in professional and academic contexts.

Let us now consider the main characteristics and the rationale for using the COCA corpus in writing instruction and the development of writing competence:

Corpus representativeness allows for the study and analysis of the entire spectrum of linguistic phenomena, enabling students to learn, recognize, and correctly use various language genres according to the communicative situation;

1. **Corpus authenticity** is “presented as a computerized database of real language, existing in various contexts, which can be used for linguistic research. Corpora can demonstrate language nuances and serve as a basis for educational materials and tasks in the language classroom”, as well as provide systematic access to texts in “their natural contextual form”, and promote not only research-based learning but also “autonomous learning and teaching” (Boulton, 2021; 563-580).

2. **Sampling criteria** serve as the foundation for the collection and classification of the corpus, meaning that text types, quantity of texts, text samples, and sample lengths can be reused by various researchers for different purposes (Radjabova, 2021; 157-163).

From our perspective, authenticity as one of the key characteristics of a corpus is particularly significant for broad application in the improvement of students' writing competence. It allows learners to access “real language and language situations, which expands students' abilities to use their language skills” (Chisman, 2008; 4-6), which in turn serves as a stimulus for the development and enhancement of writing competence. Many researchers also agree that “the use of authenticity in foreign language instruction is effective”. Gilmore (2007) defines an “authentic text” as “a piece of real language produced by a real speaker or writer for a real audience and designed to convey a real message” (Gilmore, 2007; 97-100). In his study, the author illustrates the so-called “inadequacy” of many modern textbooks, which “due to their lack of authenticity cannot help develop students' writing competence” (Radjabova, 2023). From this, it follows that providing “real, authentic English in a meaningful form is a dilemma faced by most language programs”. However, corpora and corpus technologies have made a significant leap in resolving this dilemma, allowing the inclusion of “real, authentic English” in curricula. In other words, the development of corpus technologies has paved the way for the implementation of the data-driven learning (DDL) approach in language teaching, where language use, as opposed to structure, is dominant.

3. According to Biber, Conrad, and Reppen (1998), the corpus approach allows not only the study of linguistic features and their characteristics, but also provides broad opportunities to explore how “speakers and writers use the resources of their language, meaning we study real language used in naturally occurring texts” (Biber, D., Conrad, S., & Reppen, R., 1998, p. 3). In their book, the authors emphasize that the main characteristics of the DDL approach are:

- “empirical analysis of actual language use patterns in natural texts;
- use of large collections of natural texts, known as ‘corpora’, as the basis for analysis;
- extensive use of computers for analysis, applying both automatic and interactive methods;
- direct dependence on both quantitative and qualitative analysis methods” (Biber, Conrad & Reppen, 1998, 4-5).

In other words, the DDL approach allows for in-depth analysis of authentic language, providing empirically verified data that can be used to answer language-related questions instead of relying on intuition. This approach helps to build learners' active vocabulary, thereby developing and enhancing writing competence as a component of communicative competence.

Corpus linguistics serves as the theoretical and instrumental foundation, providing methods and technologies. DDL applies these tools to create educational materials, exercises, and methodologies that help students learn language based on real data. Figure 1.6 illustrates how corpus linguistics (science and technology) is integrated into language teaching practice (DDL), creating a basis for innovative, data-driven teaching approaches. In other words, corpus linguistics provides tools and resources, such as large text collections (corpora) and specialized technologies (KWIC, CHART, TEXT ANALYSIS). These tools help to study language in its natural environment, identify word frequency, typical usage contexts, and grammatical and lexical patterns. The DDL approach takes these data and turns them into learning material. Emphasis is placed on enabling students to "discover" language rules themselves through text and example analysis. This approach fosters language intuition, analytical thinking, and independent learning skills.

References:

1. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York, NY: Cambridge University Press.
2. Boulton, A. (2021) Data-driven learning: the perpetual enigma. In: Goźdz-Roszkowski S (ed) *Explorations across languages and corpora*. Peter Lang, Frankfurt, pp. 563–580
3. Chisman, F. (2008). Findings in ESL: A quick reference to findings of CAAL research on ESL programs at community colleges. Results from the Council for Advancement of Adult Literacy. Retrieved from: <http://caalusa.org>., pp. 4 – 6.
4. Davies, M. (2009). The 385+million-word corpus of contemporary American English: Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14, 159-190. <http://dx.doi.org/10.1075/ijcl.14.2.02dav> https://www.english-corpora.org/coca/help/coca2020_overview.pdf
5. Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching* 40, pp. 97 – 118. doi: 10.1017/S0261444807004144.
6. Giyosiddinovna, Radjabova G. "Methodological Characteristics of Corpus Technologies in Teaching Foreign Language." *International Journal on Integrated Education*, vol. 5, no. 1, 2022, pp. 157-163, doi:[10.31149/ijie.v5i1.2645](https://doi.org/10.31149/ijie.v5i1.2645).
7. Radjabova, G. (2023). Corpus technologies in teaching academic writing. *Foreign Languages in Uzbekistan*, 1(48), 92-103.
8. Radjabova, G. G. (2024). ADJUSTING THE PERSPECTIVE OF CORPUS LINGUISTICS: BRIDGING RESEARCH AND THE CLASSROOM. *American Journal of Modern World Sciences*, 1(5), 324-332.
9. Нагель О. В. Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // *Язык и культура*. 2008. № 4. С. 58.