

USING CORPUS OF CONTEMPORARY AMERICAN ENGLISH [COCA] TO ENHANCE WRITING COMPETENCE OF B2 LEVEL STUDENTS

Gulnoza Giyosiddinovna RADJABOVA

Associate Professor (PhD)

Uzbekistan State World Languages University

Tashkent, Uzbekistan

B2 DARAJASIDAGI TALABALARNING YOZISH KOMPETENSIYASINI OSHIRISH UCHUN ZAMONAVIY AMERIKA INGLIZ TILI KORPUSIDAN [COCA] FOYDALANISH

Gulnoza G'iyosiddinovna RADJABOVA

Dotsent (PhD)

O'zbekiston davlat jahon tillari universiteti

Toshkent, O'zbekiston

ИСПОЛЬЗОВАНИЕ КОРПУСА СОВРЕМЕННОГО АМЕРИКАНСКОГО АНГЛИЙСКОГО ЯЗЫКА (СОСА) ДЛЯ ПОВЫШЕНИЯ ПИСЬМЕННОЙ ГРАМОТНОСТИ УЧАЩИХСЯ УРОВНЯ В2

Гульноза Гиёсиддиновна РАДЖАБОВА

Доцент (PhD)

Узбекский государственный университет мировых языков

Ташкент, Узбекистан rad.gulnoza@gmail.com

For citation (iqtibos keltirish uchun, для цитирования):

Radjabova G.G. Using Corpus of Contemporary American English [COCA] to Enhance Writing Competence of B2 Level Students // O'zbekistonda xorijiy tillar. — 2025. — 11-jild, № 6. — B. 110-124.

<https://doi.org/10.36078/1767689405>

Received: October 05, 2025

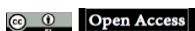
Accepted: December 17, 2025

Published: December 20, 2025

Copyright © 2025 by author(s).

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract. The rapid development of computers and Internet technologies has significantly transformed education, especially methods of teaching foreign languages. Modern language pedagogy increasingly incorporates digital tools that enhance learner autonomy and improve the quality of instruction. Among these innovative tools, corpus technologies stand out as an effective resource for developing language competence. The article examines the evolution of corpus linguistics as a scientific field and analyses its pedagogical potential in teaching foreign languages. The novelty of the study lies in its systematic integration of corpus-based tools such as concordances, corpus statistics, and text-analysis applications into a structured set of exercises for improving writing competence. The purpose of the research is to evaluate the effectiveness of corpus technologies in fostering communicative competence, particularly academic writing skills of B2-level learners. The experiment was conducted among third-year students of Uzbekistan State World Languages University and employed both quantitative and qualitative data collection methods. The results demonstrate that corpus-informed instruction contributes to increased lexical awareness, improved accuracy, and deeper understanding of semantic prosody. Furthermore, the assignments designed on the basis of the Corpus of Contemporary American English (COCA) can also be used by students independently as mini-research projects. Overall, the study confirms that corpus technologies provide rich, authentic data and significantly enhance the writing competence of language learners.

Keywords: corpus linguistics; corpus technology; concordance; semantic prosody; lexical density.

Annotatsiya. Kompyuter va internet texnologiyalarining jadal rivojlanishi ta'lim jarayonini, xususan, chet tillarini o'qitish metodlarini sezilarli darajada o'zgartirdi. Zamonaviy o'qitish jarayonida raqamli vositalardan samarali foydalanish o'quvchilarning mustaqilligini oshiradi va til o'rganish samaradorligini kuchaytiradi. Shunday innovatsion vositalardan biri korpus texnologiyalaridir. Mazkur maqolada korpus lingvistikasi fanining shakllanishi, uning ta'limdagi qo'llanilishiga oid ilmiy adabiyotlar tahlil qilinadi. Tadqiqotning yangiligi shundan iboratki, unda korpus texnologiyalari, konkordanslar, statistik ma'lumotlar, matn tahlili vositalari, yozma nutq kompetensiyasini rivojlantirishga yo'naltirilgan mashqlar tizimiga kompleks tarzda integratsiya qilingan. Tadqiqotning asosiy maqsadi korpus texnologiyalarining B2 darajasidagi talabalar yozuv kompetensiyasini rivojlantirishdagi samaradorligini aniqlashdan iborat. Eksperiment O'zbekiston davlat jahon tillari universitetining 3-bosqich talabalari o'rtasida o'tkazildi va unda miqdoriy hamda sifat tahlili metodlari qo'llandi. Natijalar shuni ko'rsatdiki, korpus asosidagi topshiriqlar leksik boylikni oshirish, semantik prosodiyani chuqurroq tushunish va yozuv aniqligini yaxshilashga yordam beradi. COCA korpusidan olingan topshiriqlar talabalarning mustaqil mini-tadqiqot loyihalari sifatida ham qo'llanishi mumkin. Maqola korpus texnologiyalarining chet tilini o'qitishda yuqori samaradorlikka ega ekanini tasdiqlaydi.

Kalit so'zlar: korpus lingvistikasi; korpus texnologiyasi; konkordans; semantik prosodiya; leksik zichlik.

Аннотация. Интенсивное развитие компьютерных и интернет-технологий существенно изменило систему образования и методику преподавания иностранных языков. Современная лингводидактика активно внедряет цифровые инструменты, позволяющие усиливать самостоятельность обучающихся и обеспечивать доступ к аутентичному языковому материалу. Одним из наиболее перспективных направлений является использование корпуса текстов в учебном процессе. В статье представлен обзор литературы по истории и становлению корпусной лингвистики, а также раскрывается ее педагогический потенциал. Новизна исследования заключается во внедрении различных корпусных инструментов, конкордансов, статистики корпуса, программ анализа текста в систему упражнений, направленных на развитие письменной компетенции. Цель исследования — определить эффективность корпусных технологий в формировании коммуникативной компетенции, прежде всего навыков академического письма у студентов уровня B2. Эксперимент проводился среди студентов третьего курса Узбекского государственного университета мировых языков с использованием количественных и качественных методов анализа. Полученные результаты показывают, что обучение с применением корпуса способствует расширению словарного запаса, улучшению точности письма и формированию понимания семантической просодии. Упражнения, основанные на данных корпуса COCA, могут применяться студентами и в формате самостоятельных мини-исследований. Исследование подтверждает высокую эффективность корпусных технологий в обучении иностранным языкам.

Ключевые слова: корпусная лингвистика; корпусные технологии; конкорданс; семантическая просодия; лексическая плотность; преподавание английского языка.

Introduction

Nowadays, the application of corpora in teaching foreign languages is gaining popularity. It is worth mentioning that a number of scholars such as D. Biber, R. Reppen, R. Carter, I.V. Rakhmanova, N.B. Gvishiani, O.G. Gorina, N. N. Abdurakhmonova, and others claim the feasibility of integration of methodology and corpus linguistics. The emergence of corpus linguistics as a field has been significant and valuable for the development of such branches of linguistics as lexicography, grammar, stylistics, translation theory, sociolinguistics, and pragmatics [O'Keeffe, McCarthy, Carter 2007: 15-22]. One of the first areas where corpus linguistics was integrated into other branches of linguistics was dictionary compiling technologies, followed by significant transformations of educational materials [Gorina 2018: 187].

Currently, not only a dictionary compiling process, but also a selection of any coursebook materials is carried out using various corpora as a source of authentic language [Gorina 2018: 187]. Such characteristics of corpus as representativeness, authenticity, visualization of word frequency data in the form of tables, graphs and word clusters make it an effective reference system and become one of the tools of a language investigation and analysis. However, despite significant transformations and changes in a dictionary compiling process and teaching materials design by means of corpus technologies, the practice of using corpus technologies in classrooms while teaching a foreign language has not widely used and developed in Uzbekistan. Therefore, this article is aimed to introduce the ways of implementing corpus technologies, as well as to reveal the potential of corpus tools and technologies in teaching foreign languages.

We sequentially consider in our research corpus-based technologies and analyses of the experience of how corpus technologies have been applied in practice at Uzbekistan State World Languages University (USWLU) among 3rd year students with B2 level of the English language proficiency. It is worth mentioning that the tasks based on corpus technologies are variable and have significant potential for use in language teaching. A set of exercises given in the article provides better comprehension of corpus-based assignments which can be implemented in both classroom practices and independent students' work. In conclusion, the necessity developing corpus competence not only among students, but also among educators is discussed. In addition, the article recommends introducing Corpus Linguistics course into the curriculum for the 3rd year students at linguistic universities

Main part

1. Corpus technologies and tools

Currently, corpus linguistics appears to be a relatively new approach in linguistics which involves not only the empirical study of language use in "real life", but also the application of computers and electronic corpora. Originally, a "corpus" is accepted as any collection of written or spoken

texts. However, when the term is used in relation to contemporary linguistics, it tends to carry a number of connotations, where some authors emphasize its “machine-readable form, sampling, and representativeness” [Sysoev 2010: 100], “finite size and idea” [Semich 2018: 40] and that corpus represents “a standard reference for the linguistic variety” [Zolotov 2020: 40]. While linguistics is divided into many research fields depending on sets of research questions, corpus linguistics serves a different role: it offers a set of methods that can be used in the study of a large number of different research questions [Gorina 2018: 187]. This implies that “a more accurate analysis of language is feasible with corpora collected in natural contexts and with minimal experimental intervention” [Zolotov 2020: 43]. It also helps to modernize “teaching of the Uzbek language as a foreign language [or as a second language] and to obtain a high-level qualification of learners”. [Abdurakhmonova, & Urdishev 2019: 132].

Now, it is worth considering some definitions that characterize corpus linguistics as a science. According to the definition given in the dictionary entry of the new Methodological Dictionary edited by E.G. Azimov and A.N. Shchukina [2009: 114], corpus linguistics is focused on the “analysis of authentic corpora [databases] of texts”, and “on the applied study of language, its functioning in real environments and texts”. American scholar J. Sinclair [Sinclair 1991: 489], believes that “a new understanding of the nature and structure of a language will soon be available as a result of computer-assisted study of large collections of texts.” And this type of a research has been called “corpus linguistics.” According to American scholar W. Francis [Francis 1982:7], corpus linguistics is the most closely related to “the empirical analysis of electronically stored linguistic data which is natural and authentic by origin.” Corpus research seeks to create descriptions of a language based on the observation of a language in use. According to British scholar G. Kennedy [1998:5], corpus linguistics, having automatic and interactive techniques as a basic defining feature, can analyze linguistic data automatically. Russian linguists also define this discipline as a science. V. Zakharov [Zakharov 2011: 7], for instance, argues that “Corpus linguistics is a branch of computer linguistics which deals with the development of general principles of constructing and using linguistic corpora [text corpora] using computer technologies.”. However, due to insufficient development of corpus linguistics as a science, scholars are still debating about object and subject of corpus linguistics. For example, V. Mamontova [Mamontova 2010: 230] considers that the definition given by V. Zakharov is just a narrow understanding of the discipline which limits it to the scope of computer linguistics while “computer linguistics is usually understood as a broad field of computer tools, computer technologies for organizing and processing data, to simulate a language functioning in particular conditions, situations, problem areas.” V. Mamontova [Mamontova 2010: 230] believes that corpus linguistics uses computers as its tool but without computers, of course, it would not be able to perform its functions. Thus, to explore the analytical potential of language, this article examines several tools and technologies employed in corpus linguistics.

1.1 Concordance

The generation of concordance lines is one of the fundamental tools used in corpus linguistics. The concordance lines are presented in the form of “corpora of text extracts with the lexical unit under consideration” [Gorina 2014: 13]. T. Johnes, one of the first scholars to suggest using concordance lines in language teaching, argues for a “discovery learning” approach in which learners “discover” the foreign language while teachers create the conditions that enable such discoveries. [Johns 1991: 1] Thus, a learning model based on the linguistic visualization of the concordance was proposed. A learner gets access to the language patterns used by the speech community and at the same time acts as a researcher [Johnes 1991: 6]. The approach of vertical reading or scanning of concordance creates a lot of possibilities to disclose “the most likely grammatical structure of the word and its typical combination with other lexical units” [O’Keeffe, McCarthy, Carter 2007: 9]. In other words, concordance enables learners to notice important features of lexical use in “the most probable grammatical environment” and study independently the compatibility of words [O’Keeffe, McCarthy, Carter 2007: 15]

Studying compatibility of lexical units helps students to prevent errors and interference. N. B. Gvishiani [Gvishiani 2008: 12] notes that contrastive analysis of the speech production of the language learners reveals that most errors are caused by the interference of L1 and a foreign language. According to data taken from essays of the third-year students of USWLU, a specific mistake of the students whose L1 is either Russian or Uzbek is the use of the phrases **“discuss something/discuss about something: enter somewhere/enter to somewhere”**. COCA data show that **“discuss something”** occurs approximately 47800 times and **“enter somewhere”** about 47000 cases, while the option of **“discuss about something”** is given in 85 cases and mostly in the spoken register. Similarly, **“enter to”** appears only 45 times, and there are no recorded instances of the phrase **“enter to somewhere”**. Thus, by studying 10-15 concordance lines, learners can independently draw the conclusion that the expressions **“discuss about something”** and **“enter to somewhere”** should “be avoided” [Gvishiani 2008: 39].

1.2 The extended context in the corpus

Another corpus tool is the extended context option, which allows the learners to examine the full context in which the word under analysis occurs “The extended context helps maintain contextual completeness by giving information about the temporal and situational background of the text, its author and its publication source” [Gorina 2014: 188].

However, some researchers claim that a corpus, especially a large, diversified corpus [e.g. BNC – The British National Corpus: COCA – The Corpus of Contemporary American English:] is a “museum” of texts where inclusion is lost because “the text does not accompany the reality” [Brawn 2007: 32-33]. But the extended context helps to present a real situation indicating how a word has been used in different contexts for a long time. In addition, the implementation of a small, specialized corpus has a number of advantages because its usage eliminates the problems connected with inclusion [Gorina 2013: 202]. One of the advantages of

“small, specialized corpora” is that it can be compiled by a team of teachers and it can be oriented towards the needs of students of a particular university [Radjabova 2023: 34].

Since no single corpus can effectively serve all purposes at once, the creation of specialized university-based so called “home-made” corpora is gaining popularity and becoming a common practice. The goal of such university projects is to compile professionally oriented linguistic databases, such as a corpus of teachers’ discourse, a corpus of learners’ errors, a corpus of different academic disciplines [Fil’chenko & Shalamova 2004: 23].

1.3 Corpus statistics visualization and semantic prosody

The main methods of working with a corpus involve various types of linguistic statistics, such as indentifying word frequency, determining the most frequent uses of lexical units in different genres [including professionally relevant ones], defining key words and word-clusters through corpus searches for word combinations [word collocations which may be used to search words in corpus]. [Gorina 2018: 189] These corpus functions rely on statistical visualization of large linguistic datasets. But the validity and accuracy of corpus-based findings still depend on the intuition of native speakers. Therefore, non-native speakers have the opportunity to evaluate the authenticity of written and spoken language through linguo-statistical indicators [Klepikova & Klushkina 2012: 285].

Semantic prosody, for example, describes the typical contextual associations whether positive or negative [Gvishiani 2008: 24]. For example, the verb **cause** [“to be a reason”] has negative connotations with such lexical units as **cancer, crisis, delay** [Stubbs 1995: 46]. This verb has negative prosody in about 80% cases of its usage and only about 20% of collocations are positively connotated. If we compare the percentage of negative and positive connotations of this verb, there is a clear tendency of combining this verb with negative words: 80 vs. 20%. This tendency can be explained by the semantic structure of the verb **cause** which denotes bringing something about or prompting an event or state. [Stubbs 1995: 48]. Corpus-based statistical measures have significant explanatory potential and can be used to assess the authenticity of an utterance by the educator and set clear goals of the research. In contrast to the verb **cause**, the verb **provide** tends to have positive connotations in collocations with the words **care, help, relief, support** [Stubbs 1995: 49]. The phenomenon has not been studied in “pre-corpus” era, and there was no possibility of quantitative confirmation of authentic materials.

1.3.1 Corpus search for collocations

When discussing collocation searches, it is necessary to point out some of the features of corpus tools. If the automatic search for collocations is done, both the word under analysis and collocational environment are displayed. For example, the search can be set to display one to five words to the left or right of the target word. In addition users can specify genre, register and domain of knowledge.

While searching for words which can collocate “with a noun **stigma**, the most frequent phrase **a huge stigma**” is shown in key word in the

context [KWIC] research [Gorina 2018: 189]. If we expand the search to 2-4 words which can be combined with the target word **stigma**, the so-called “chunk [the formation of a typical collocations] like **a huge stigma attached to...**” can be found. It is worth mentioning that the ability to operate with such chunks constitutes the basis of fluent speech of native speakers. [Gorina 2018: 189] Therefore, identifying of such statistical features of speech is an important professionally relevant skill for those who teach a foreign language. This skill is especially important for L2 learners because it promotes the formation of correct compatibility. Corpus-based examples prevent interference if a speaker has little or even no experience in collocation usage [Gorina 2013: 202], because there are cases of “deformation of collocations” such as “**his eggs were in too many baskets**” [original phrase is “**don’t put all your eggs in one basket**”] that are difficult to use by speakers whose native language is not English [Ter-Minasova 2009: 31]. Depending on the aim of communicative act such as persuasion or confirmation, such deformations may occur. Therefore, learners can learn how to investigate corpus in terms of “distribution of idiomatic expressions in different genres of the corpus” and use them “by selecting only relevant expressions in the context of their study” [Gorina 2014: 190]. Such corpus analysis develops not only research skills, but also autonomous learning skills of learners.

1.3.2 Lexical density

The correlation between the number of word forms and the number of lexemes [type/token ratio] is also considered to be one of the key functions of corpus technologies. The indicator of such correlations is called “lexical density” [Gvishiani 2010: 48]. Certain peculiarities and register-specific features of written and spoken foreign language texts can be revealed by means of the lexical density. According to corpus-based investigations of Biber D., Johansson S., Leech G. [1999: 62], in terms of English Grammar, everyday oral conversations are less lexically dense than news reports. While news articles are carefully edited and proofread, the everyday dialogues between people occur in “live” format when there is no possibility to edit grammatical errors and inaccuracies, the correction of what is said might occur in the following utterances and the utterance itself cannot be deleted. The objects surrounding the interlocutors can be mentioned and involved into the situation by the inclusion of gestures or such words as “**erm, right, yeah, please, gosh, well**”. On average, in LSWE [Longman Corpus of Spoken and Written English], for example, lexical [meaningful] words constitute almost half of all words in this corpus. However, the proportion of functional words in the written discourse is higher than in spoken discourse, about 44% and 37%, respectively, and the inclusion of such words as **erm, right, yeah, please, gosh** is observed only in spoken discourse – 15% [Biber, Johansson, Leech, et al., 1999: 65].

Such corpus statistics have become an essential part of grammar textbooks, dictionaries and other teaching materials created for learners of a foreign language because they are significant in the process of development of professional competence of future language specialists. Furthermore, visualization of corpus-based statistics is also important for

translation studies [the texts of the originals and their translations], as close attention can be paid to the peculiarities of language equivalences and analogues. In other words, statistical visualization serves as a reliable basis or source in researching the translation challenges while comparing original texts and their translation [Tognini-Bonelli 2002: 67].

2. Corpus-driven tasks and exercises

In this section, a number of tasks as well as some small independent research projects will be presented. The materials are taken from the database of Corpus of Contemporary American English [COCA].

2.1 Example of a concordance-based assignment

This task is aimed at mastering lexical units, which are difficult to use in everyday communication and this task is recommended to students with B2 proficiency level. However, the concordance sorting tasks can be used at earlier stages of language learning: namely, it can be applied not only for B2 level learners, but also for A2 and B1. \ The applicability of the task depends on the corpus-literacy and computer skills of learners and desire of an educator to develop those skills. In this task learners are asked to transform the information presented in concordance into a so-called “collocation profile” of the word. The task is aimed at the development of collocation competence of learners. In the following table the fragment of the noun **resilience** is presented [Tab. 1].

Table 1

Collocation profile of the noun resilience

blossoms, defying all odds and overcoming difficulties through the resilience and dedication of the Ogaden people's selfless sacrifices and heroism.
judy415: “Shauna, I admire your enormous resilience of character and personal resourcefulness in becoming an attorney and raising your daughter as
resilient seeds for poor farmers in Africa. Work on resilience also enabled the development of early-warning systems in Nepal during the 2010 floods
only remains now. The grounds were beautiful reflecting the resilience of these women. # There were no trailers in Ishinomaki but the temporary
new policy to help vulnerable communities across the world build resilience to future crises. # The new Communication, “The EU Approach to
new policy to help vulnerable communities across the world build resilience to future crises. # The new Communication, “The EU Approach to
and “agriculture, “advocates a sort of soft resilience to withstand the forces of nature, not of brick or reinforced concrete but

The expected result is that students will note collocations which are

the most frequent in this full concordance with the adjectives **enormous, a sort of, important, great, incredible, seismic** and will mark the other most important grammatical features of the word, namely, grammatically possible variants of the noun with the verbs **build** and **promote** in pre-position.

In terms of grammatical variants of the noun **resilience**, there is a clear tendency towards the combination with the verbs in pre-position, while the combination of the verbs in post-position is rarely observed. There is also a clear tendency to make a collocation scheme with the verb **to build** [42 occurrences out of 264 in the COCA], and less frequently with the verb **to promote and create**. A noticeable syntactic structure for the combination **to build resilience** is that it is mostly used in the function of an adverbial modifier of purpose [Table 2].

Table 2

Combinability and frequency data of the phrase to build resilience

#to be raised when the economy is healthy to build resilience against losses and relaxed during a downturn to encourage greater lending. # Officials
#they told themselves they could do it. To build resilience, praise effort and emphasize that the child simply has not mastered a skill
#used to say the goal of CVE was to build resilience within the communities, but it was actually quite divisive within the community.
#them where they are leaking time and how to build resilience to that. # First of all, you should look inside your company
#to be. What I have to do is to build resilience into Barclays so that we can cope with those things, and where possible
#their expectations, “Fredrickson says. You can build resilience with a steady diet of positive emotional experiences. That can be as simple
#classrooms operated by the Allegheny Intermediate Unit help students build resilience is by following a daily routine. “Children like to know they are
#and they do not necessarily help an individual to build resilience, to overcome difficulties as they happen or to react to challenges with composure
#needs “with early intervention and work “to build resilience prior to combat deployments. “That is not true. “Navy medicine
#. So there are multiple ways that you can build resilience in a way that actually has, you know, physiological correlates that you

Further tasks which students can do to consolidate certain lexical units is to make their own sentences using those collocations of the given lexical units, search for examples of them in mass media articles, and implement these units in communication tasks.

2.2 “Problem solution” tasks

Our research has shown that corpus statistics regarding the combinability of words and their usage in different registers and contexts has a significant explanatory potential. This research assignment aims to

investigate and analyze register/genre peculiarities of lexical units.

The instruction of the assignment is as following:

Using the genre breakdown of the COCA, investigate the usage of the verbs “exclude” and “rule out” and find why these two synonymous verbs are used differently. The possible solution of this task is that according to the data from the corpus the collocation “**rule out the possibility**” occurs mostly in news articles and newspaper texts, while the collocation “**exclude the possibility**” is common only in academic texts. Thus, as it is seen from the examples above, the usage of the verb “exclude” in non-academic context does not meet the expectations of native speakers as it sounds unnatural to them.

2.3 Sociolinguistic variable analysis task

Sociolinguistic variables are accepted as analytical tools when “on different occasions and for different purposes, English is used as a means of communication” [Gvishiani 2008: 53]. In such tasks learners can be asked to analyze how social variables such as gender and age are reflected by language means to express ‘*agreement/disagreement*’, ‘*certainty/uncertainty*’, ‘*disappointment*’, ‘*annoyance and anger*’, ‘*permission*’, ‘*requests*’, and other meanings associated with Social English. The corpus data can be used to demonstrate how social variation and language use are related [Gvishiani 2008: 53]. The expressions associated with vague language [*sort of, not quite, somewhat, I suppose, etc.*] can also be analyzed with reference to speaker’s age as an important parameter of social differentiation in language use [Tab. 3, 4].

Table 3

Transcription of the corpus-based spoken dialogue

[SP:PS0FF]	She said, they hadn't! She said!
[SP:PS0FN]	And they're in.
[SP:PS0FF]	We haven't had anybody, have we?
[SP:PS0FM]	Mum, will daddy be here? [pause] Why doesn't daddy come?
[pause]	Granddad.
[SP:PS0FN]	He'll bloody come next week. Next time, I suppose.
[SP:PS0FM]	Why?
[SP:PS0FN]	Cos I'll make him!
[SP:PS0FM]	But it's somewhat strange, isn't it?
[SP:PS0FF]	We let him out.
[SP:PS0FN]	No, I've been meaning to.
[SP:PS0FM]	Come on, Mark!
[SP:PS0FN]	See you later!
[SP:PS0FF]	Mark'll go bananas, I expect!
[SP:PS0FN]	Well, now how long have they known you were going to rent a caravan?
[SP:PS0FG]	Well, it's been booked for three or four weeks, hasn't it?
[SP:PS0FF]	Ooh, more than that!

Table 4*Speaker's background information*

Speaker information for PSOFF	
Name:	Josephine
Number of words:	19,651
Number of turns:	2,039
Standard header information:	
Sex:	Female
Age:	45-59
Social Class:	C1
Education:	n/a
First Language:	n/a
Dialect/Accent:	Welsh
Role:	housewife

According to the instruction, students can be asked to compare formal/informal, academic/non-academic usage of lexical units paying attention to differences and similarities in \ vocabulary use and certain collocations. For example, whether the collocation *“go bananas”* can only be used in informal texts or there are some instances of its implementation in academic contexts. Additionally, learners can be asked to analyze how formulaic expressions like *“well”*, *“I mean”*, *“you know”*, *“you see”* and reaction signals are differentiated by the parameter of a speaker's age. The corpus-based spoken dialogues seem to be more reliable than those found in textbooks and dictionaries as they provide and add to learners' background knowledge more facts about the speakers, namely, their age, gender, level of education, social position, the nature of relationship between them [see Table 4]. Besides such corpus texts give information about the time and place of the speech event and, the topic of the conversation. The style and register features accompanying a given speech act, such as dialogue/monologue, informational/creative and others are also of great importance. These types of information present a complete discourse unit, which merges structural, semantic, and contextual dimensions [Tognini-Bonelly 2002: 79].

2.4 Text analysis task

There are numerous problems connected with word usage which learners might face while learning a language. In actual speech, different word forms usually adhere to certain contexts with varying degrees of regularity. In this case the frequency of word-forms based on such features as their functional capacity, range of collocability, and utility can be differentiated. The most challenge is that some forms of words possess greater discourse potential and are used more frequently in certain types of discourse than others. The adjective *usual*, for example, is less frequent than the adverb *usually* in both Written and Spoken English according to COCA (about 43000 occurrences of *usual* in comparison to 123000

occurrences of *usually*). Therefore, one of corpus tools which may be beneficial for learners to comprehend the frequency of a word-form usage is connected with the opportunity to analyze the text in a special text engine offered by COCA. Learners can enter any text, for example, a paper that they have written, or a newspaper article that they have copied from another website. After inserting the text, learners are provided with the information based on the data from COCA about the word frequency visually highlighted in different colors. This **frequency data** can help language learners focus on new words, and it can enable to see “what the text is about” (i.e. text-specific words).

As a part of the task, each student can be assigned to find a pair of the most and the least frequently used words/phrases from the text under the analysis in COCA. They will have to explore each word/phrase in terms of the differences in usage, location in the sentence, punctuation, and the part of speech of the word that comes after or/and before the search word. In addition, students can be asked to post their findings, or create their own texts either oral or written using the least frequently used words/phrases. In this situation, the context reveals its connotations or ‘preferences’ in terms of communication and discourse. Corpora studies of a text show us how frequently the word is used and how regularly the word associates with other meanings. A text analysis gives information not only about typical associations of a particular word, but also about other tendencies in word-grouping. [Gvishiani 2008: 94]

Results of the research

Corpus technologies having not only such characteristics as empirical relevance, authenticity, representativeness, flexibility and adaptation regarding specific tasks, but also the opportunity for students to work autonomously, can produce:

- 1) the usage of a chosen word in its immediate context, on the basis of which students can analyze due to the aim of their research or task:
- 2) a key word in the context (KWIC) as a tool which makes it possible to investigate the combinability features of verbs, nouns, etc.:
- 3) intertextuality that reveals lexico-semantic variants of a word as the sum of its uses:
- 4) a text analysis tool to develop writing competence of learners with regards to the usage of high, middle and low frequency words:

The assignments based on the usage of corpus statistical visualization can be combined with other traditional tasks aimed to consolidate vocabulary of the language. Besides, corpus-based learning techniques and tools can be combined with communicative tasks, i.e., the learners can apply language units in practice. But the validity of the suggested assignments is confirmed by the results of the experimental teaching conducted on writing among 3rd year students in USWLU. According to the results of the experiment aimed to develop writing competence and increase vocabulary skills of learners, the overall grade in the experimental groups increased by 4-5% in comparison to the control groups. Besides, the development of a number of skills, such as independent learning, problem-solving abilities, and a higher degree of students' involvement into the learning process have been observed.

It should be admitted that in order to take advantage of the opportunities which corpus can provide, both students and educators need to become familiar with the rules corpus query and master the rudimentary skills of working with a corpus database. Consistent investigations of corpus linguistics application in linguo-didactics enable to demonstrate convincingly its potential for developing the key foreign language competencies. The implementation of corpus tools enables both teachers and students to develop skills of an autonomous researcher by having the access to a large, authentic reference system.

Conclusion

Therefore, the development of corpus competence among educators and students, which can be accomplished by gaining access to corpus-based data and corpus technologies as a source of authentic materials, will not only improve and develop the conventional methods of teaching foreign languages that are currently in use, but it will also bring “nativeness” into the classroom. Corpus technologies, which have characteristics such as authenticity, representation of different registers, linguistic and statistical visibility, and a variety of search capabilities presented in a linguistic and methodological context, are able to effectively develop all language competencies because of their reliance on the considerable potential of these technologies. On top of that, corpus data “can assist in the creation of authentic situations,” and “lessons involving authentic” data “can be featured to increase exposure to the language associated with each.” In the research that was shown earlier, the activities that were emphasized are examples of the kinds of innovative techniques that can be implemented into instructional materials. It is possible for the ideas to be presented to the teachers, which means that they will not be required to search for and develop materials and activities [Radjabova 2021: 351]. We have presented an outline of the primary directions that the development of language competence is heading in through the research that we have conducted. Not only does the experience of assembling a small, professionally-oriented corpus based on instructional materials and given in the article above indicate the collaboration of a team of teachers from one department, but it also suggests the collaboration of specialists from other disciplines. When determining whether or not it is possible to compile a corpus of a certain educational institution, the priority is placed on picking the materials that are the most pertinent and precise to be researched.

References

- Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics*, 6(1), 131–137.
- Azimov, E. G., & Shchukin, A. N. (2009). *New dictionary of methodological terms and concepts (Theory and practice of language teaching)*. IKAR.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. Pearson Education Limited.
- Brawn, S. (2007). Designing and exploiting small multimedia corpora for autonomous learning and teaching. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom*. Rodopi.
- Francis, W. N. (1992). Language corpora B.C. In J. Svartvik (Ed.), *Directions in corpus linguistics*. Mouton de Gruyter.
- Gorina, O. G. (2013). Using corpus evidence to compile learner's vocabulary. *Vestnik LGU im. A. S. Pushkina*, 7(2), 201–209.
- Gorina, O. G. (2014). Using the technology of corpus linguistics for the development of lexical skills of students majoring in regional studies (PhD thesis). Moscow.
- Gorina, O. G. (2014). Psychological grounds of using corpora in the ESP classroom. *Vestnik LGU im. A. S. Pushkina*, 7(1), 172–179.
- Gorina, O. G. (2018). Tools of corpus analysis in teaching foreign language. *Tomsk State University Journal*, 435, 187–194. <https://doi.org/10.17223/15617793/435/24>
- Gvishiani, N. B. (2008). Textbook on corpus linguistics: Studies. An English manual. Vysshaya Shkola.
- Gvishiani, N. B. (2010). Introduction to contrastive lexicology (English–Russian interlanguage correspondence). Librokom.
- Johns, T. (1991). Should you be persuaded – two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal*, 1–16.
- Kamshilova, O. N., Kolina, M. V., & Nikolaeva, E. A. (2008). Development of a corpus of texts of Petersburg schoolchildren: Tasks and prospects. In *Applied linguistics in science and education* (pp. 92–98). St. Petersburg.
- Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 95–110). Longman.
- Klepikova, T. A., & Klushina, Zh. V. (2012). The study of terminological collocations from the perspective of cognitive grammar and corpus linguistics. *TSU Bulletin*, 12, 284–288.
- Mamontova, V. V. (2010). Corpus linguistics in modern paradigm. *Actual Issues of Modern Science*, 12, 230–238.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Radjabova, G. G. (2021). The implementation of spoken corpora in creating teaching materials. *International Journal on Integrated Education*, 4(5), 349–354.

- Radjabova, G. G. (2022). Corpus-based approach in teaching foreign languages. *Conferencea*, 100–103. <https://www.conferencea.org/index.php/conferences/article/view/1236>
- Radjabova, G. G. (2022). Methodological characteristics of corpus technologies in teaching foreign language. *International Journal on Integrated Education*, 5(1), 157–163. <https://doi.org/10.31149/ijie.v5i1.2645>
- Radjabova, G. G. (2023). Corpus technologies in teaching academic writing. *O'zbekistonda xorijiy tillar*, 1(48), 92–103. <https://doi.org/10.36078/1679549918>
- Radjabova, G. G. (2023). The integration of corpus linguistics and language teaching. *International Scientific Conferences with Higher Educational Institutions*, 3(08.05).
- Radjabova, G. G. (2024). Features of written corpora exploited in creating language teaching materials. *Xalqaro ilmiy-amaliy konferensiyalar*, 1(1), 172–176.
- Rakhmanov, I. V., et al. (1972). The main directions in the methodology of teaching foreign languages in the nineteenth and twentieth centuries. *Pedagogika*.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Semich, Yu. I. (2018). Didactic properties and methodological functions of corpus technologies. *Bulletin of Tambov University. Series: Humanities*, 23(177), 39–49. <https://doi.org/10.20310/1810-0201-2018-23-177-39-49>
- Shalamova, N. N., & Fil'chenko, A. Yu. (2004). Corpus linguistics and its use in the profile-oriented teaching of foreign languages. *TPU*.
- Sinclair, J. (1991). Shared knowledge. In J. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics*. Georgetown University Press.
- Stubbs, M. (1995). Corpus evidence for norms of lexical collocation. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics*. Oxford University Press.
- Sysoev, P. V. (2010). Linguistic corpus in foreign language teaching methodology. *Language and Culture*, 1, 99–111.
- Ter-Minasova, S. G. (2009). *Speech syntagmatics: Ontology and heuristics* (2nd ed.). Librokom.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. In B. Altenberg & S. Granger (Eds.), *Lexis in contrast*. John Benjamins.
- Zakharov, V. P. (2011). *Corpus linguistics: Textbook for students of humanitarian universities*. IGLU.