



Received: January 27, 2022
Accepted: March 22, 2022
Available online: March 25, 2022

Gulnoza Radjabova

Senior Teacher, Uzbekistan State World
Languages University
Tashkent, Uzbekistan
E-mail: rad.gulnoza@gmail.com

Kamola Muradkasimova

DSc Student, Uzbekistan State World
Languages University, Doctor of Philosophy in
Pedagogical Sciences
Tashkent, Uzbekistan
E-mail: kmuradkasimova@yahoo.com

THE USAGE OF CORPUS TOOLS AND TECHNOLOGIES IN TEACHING AND ASSESSING ENGLISH

ANNOTATION

With the rapid development of the computers and the Internet, the education in general and the process of teaching in particular have undergone considerable changes. Teaching methods and techniques have been adapted to modern, computer-based realities. So far, a lot of techniques in terms of teaching languages with the help of computers and the Internet have been discovered. One of such techniques is considered to be the use of different corpora in teaching foreign languages.

This article overviews the literature on the development of corpus linguistics as a science and its implementation in teaching languages. It also identifies different corpus tools and technologies and their possible application in teaching foreign languages. In addition, the effectiveness and teaching potential of corpus technologies in the development of communicative competence, namely, writing competence, as well as the ways of integrating corpus technologies into the learning process are described. Finally, examples of assignments based on corpus data taken from the Corpus of Contemporary American English (COCA) are provided in this article. The experiment has been carried out among the 3rd year students of Uzbekistan State World Languages University (USWLU) using quantitative and qualitative data collection tools. Such corpus technologies as concordances, corpus

Gulnoza Radjabova

O'zbekiston davlat jahon tillari universiteti
katta o'qituvchisi
Toshkent, O'zbekiston

Kamola Muradkasimova

O'zbekiston davlat jahon tillari
universiteti doktoranti
Pedagogika fanlari bo'yicha falsafa doktori
Toshkent, O'zbekiston

INGLIZ TILINI O'QITISH VA BAHOLASHDA KORPUS VOSITALARI VA TEXNOLOGIYALARIDAN FOYDALANISH

ANNOTASIYA

Kompyuter texnologiyasi va internetning jadal su'ratlarda rivojlanishi tufayli umuman ta'lim, xususan, o'qitish jarayoni sezilarli o'zgarishlarga duch keldi. O'qitishning uslub va usullari zamonaviy, kompyuter asosidagi voqelikka moslashtirildi. Hozirgacha kompyuter va internet yordamida tillarni o'rgatish nuqtai nazaridan juda ko'p usullar kashf etilgan. Bunday usullardan biri – chet tillarni o'qitishda turli korpuslardan foydalanishdir.

Ushbu maqolada korpus lingvistikasining fan sifatida rivojlanishi va uning tillarni o'qitishda tatbiq etilishiga oid masalasi ko'rib chiqilgan. Shuningdek, turli korpus vositalari va texnologiyalarini hamda ularning chet tillarni o'rgatishda qo'llanilishi mumkinligi aniqlanib, tahlil qilingan. Bundan tashqari, kommunikativ kompetensiyani, ya'ni yozish kompetensiyasini rivojlantirishda korpus texnologiyalarining samaradorligi ko'rsatib berilgan va talabalarning o'quv salohiyati va korpus texnologiyalarini o'quv jarayoniga integratsiyalash usullari tavsiflangan. Nihoyat, Zamonaviy Amerika Ingliz Korpusidan (Corpus of Contemporary American English – COCA) olingan korpus ma'lumotlariga asoslangan topshiriqlar misollari keltirilgan. Tajriba O'zbekiston davlat jahon tillari universiteti (O'zDJTU) 3-kurs talabalari o'rtasida miqdoriy va sifatga oid

statistics, text analysis tool, as rich illustrative materials, were applied in elaboration of the exercises and tasks. The experiment results have shown that implementation of corpus technologies in language teaching classroom appears to be effective to foster writing competence of learners. In addition, presented teaching materials and assignments can be implemented independently by students as a mini-research project.

Key words: corpus linguistics, corpus statistics, corpus technology, concordance, collocability, semantic prosody, lexical density.

ma'lumotlarni yig'ish vositalaridan foydalan-gan holda o'tkazilgan. Mashq va topshiriqlarni ishlab chiqishda boy illyustrativ materiallar si-fatida konkordanslar, korpus statistikasi, matnni tahlil qilish vositasi kabi korpus texnologiyalaridan foydalanilgan. Eksperiment natijalari shuni ko'rsatdiki, til o'rgatish sinfida korpus texnolo-giyalarini joriy etish o'quvchilarning yozish malakasini oshirishda samarali usul hisoblanadi. Bundan tashqari, taqdim etilgan o'quv material-lari va topshiriqlar talabalar tomonidan mini-tadqiqot loyihasi sifatida mustaqil ravishda amal-ga oshirilishi mumkin.

Kalit so'zlar: korpus lingvistikasi; korpus statistikasi, korpus texnologiyasi, muvofiqlik, yig'ish qobiliyati, semantik prosodiya, leksik zichlik.

INTRODUCTION

Nowadays, the application of corpora in teaching foreign languages is gaining its 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.Abdurakhmonova, and others claim the feasibility of integration of methodology and corpus linguistics. The emergence of corpus linguistics as a science was significant and valuable in the development of such branches of linguistics as lexicography, grammar, stylistics, translation theory, sociolinguistics, and pragmatics [O'Keefe A., McCarthy M., Carter R., 2007; 15-22]. One of the first approaches of the usage and integration of corpus linguistics into other branches of linguistics has become a change in dictionary compiling technologies, followed by significant transformations of educational materials [Gorina O., 2018; 187].

Currently, not only a dictionary compiling process, but also a selection of any course book materials is fulfilled by means of different corpora as a source of authentic materials [Gorina O., 2018; 187]. Such characteristics of corpus as representativeness, authenticity, visualization of frequency of words usage 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 implementation of 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. It is worth mentioning that the tasks based on corpus technologies are variable and

have significant potential while applying in 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 and importance to develop corpus competence of not only students, but also educators will be discussed. In addition, the recommendation of the introduction of Corpus Linguistics course in the curriculum of the 3rd year students of a linguistic university is presented.

LITERATURE REVIEW

Corpus technologies and tools

Currently, corpus linguistics appears to be a relatively new approach in linguistics which involves not only the empirical study of a language use in “real life”, but also applies 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 P., 2010; 100], “finite size and idea” [Semich Y., 2018; 40] and that corpus represents “a standard reference for the linguistic variety” [Zolotov P.Y., 2020; 40]. While linguistics is divided into many research fields depending on sets of research questions, corpus linguistics has diametrically opposite functions: it offers a set of methods that can be used in the study of a large number of different research questions [Gorina O.G., 2018; 187]. It means that “a more accurate analysis of a language is feasible with corpora collected in natural contexts and with minimal experimental intervention” [Zolotov P.Y., 2020; 43]. It also helps to renovate “teaching the Uzbek language as a foreign language (or as a second language) and to obtain a high-level qualification of learners” [Abdurakhmonova N. & Urdishev K., 2019; 132].

Now, it's worth to consider 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, corpus linguistics is focused on the “analysis of authentic corpora (databases) of texts”, and “on the applied study of a language, its functioning in real environments and texts” [Azimov E., Shukin A., 2009; 114]. American scholar, J. Sinclair, 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” [Sinclair J., 1991; 489]. And this type of a research has been called “corpus linguistics”. According to American scholar, W. Francis, corpus linguistics is the most closely related to “the empirical analysis of electronically stored linguistic data which is natural and authentic by origin” [Francis W., 1982; 7]. Corpus research seeks to create descriptions of a language based on the observation of a language in use. According to British scholar, G. Kennedy, corpus linguistics, having automatic and interactive techniques as a basic defining feature, can analyze linguistic data automatically [Kennedy G., 1998; 5]. Russian linguists also define this discipline as a science. V. Zakharov, for instance, thinks 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” [Zakharov V., 2011; 7]. However, due to insufficient development of corpus linguistics as a science, scholars are still debating over object and subject of corpus linguistics. For example, V.Mamontova 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” [Mamontova V., 2010; 230]. V.Mamontova believes that corpus linguistics uses computers as its tool but without computers, of course, it would not be able to perform its functions. Thus, in order to make the analysis of language possibilities, there are some tools and technologies of corpus linguistics which are examined in this article.

Concordance

The generation of concordance lines refers to one of the tools of corpus. The concordance lines are presented in the form of “corpora of text extracts with the lexical unit under consideration” [Gorina O., 2014; 13]. T.Johnes, who was among first scientists who suggested teaching with the implementation of concordance lines, claims that “discovery learning”, when “the task of the learner is to “discover” the foreign language”, while a teacher creates conditions in which these discoveries can occur [Johns T., 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 T., 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 A., McCarthy M. & Carter R., 2007; 9]. In other words, concordance allows learners simultaneously notice important features of lexical units’ application in “the most probable grammatical environment” and study independently the compatibility of words [O’Keeffe A., McCarthy M. & Carter R., 2007; 15].

Studying compatibility of lexical units helps students to prevent errors and interference. N.B. Gvishiani 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 [Gvishiani N., 2008; 12]. So, 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”**. If we look at COCA, the most frequently used is **“discuss something”** – about 47800 cases, **“enter somewhere”** – about 47000 cases, while the option of **“discuss about something”** is given in 85 cases and mostly refers to a spoken language. Similarly, the usage of **“enter to”** is given in 45 cases, but there are no cases regarding the phrase **“enter to somewhere”**. Thus, studying 10-15 concordance lines, a learner can independently draw the conclusion that the expressions **“discuss about something”** and **“enter to**

somewhere” should “be excluded from frequent usage” [Gvishiani N., 2008; 39].

The extended context in the corpus

Another technology of the corpus tools is the extended context option, by which the learners can easily examine the whole context from which the word under analysis is originated. “The extended context helps to keep the situation inclusive by giving information about the time and circumstances of the text, the author and the source of the publication” [Gorina O., 2014; 188].

Some researchers, however, 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 S., 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 O., 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.

As it is known that 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 daily university 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 A.Y., Shalamova N.N., 2004; 23].

Corpus statistics visualization and semantic prosody

The main methods of working with a corpus include all kinds of linguistic statistics, namely, revealing the frequency of words, singling out the most frequent usage of lexical units in different genres (including professionally relevant ones), defining key words and word-clusters; corpus search for word combinations (word collocations which may be used to search words in corpus) [Gorina O., 2018; 189]. Aforementioned functions of corpus are based on statistical visualization of large linguistic data. But the appropriateness and correctness of corpus-based data is based on the intuition of native speakers. Therefore, non-native speakers have the opportunity to evaluate the authenticity of written and spoken speech with the help of linguo-statistical indicators [Klepikova T., Klushkina Z., 2012; 285].

Semantic prosody, for example, describes the probable use of a word in certain contexts, either positive or negative [Gvishiani N., 2008; 24]. For example, the verb **cause** (“to be a reason”) has negative connotations with such lexical units as **cancer, crisis, delay** [Stubbs M., 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%. Such tendency can be explained by the semantic structure of the verb **cause** as *to serve as a cause or occasion of something* or *to compel by command, authority, or force* [Stubbs

M., 1995; 48]. A variety of corpus statistics has considerable 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 M., 1995; 49]. The phenomenon has not been studied in “pre-corpus” era, and there was no possibility of quantitative confirmation of authentic materials.

Corpus search for collocations

Speaking about the collocation search, it is necessary to point out some of the features of corpora. If the automatic search for collocations is done, both the word under analysis and collocation set are displayed. For example, from 1 to 5 words, either on the left or on the right side of the word can be searched. In addition to this, genre, register and field of knowledge can be chosen.

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 O., 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 O., 2018; 189]. Therefore, in our opinion, the identification 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 O., 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 S., 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 O., 2014; 190]. Such corpus analysis develops not only research skills, but also autonomous learning skills of learners.

Lexical density

The correlation of the number of word forms and the number of lexemes (type/token ratio) is also considered to be one of the functions of corpus technologies and tools. The indicator of such correlations is called “lexical density” [Gvishiani N., 2010; 48]. Some peculiarities and register-specific features of written and spoken foreign language can be revealed by means of the lexical density. According to recent corpus-based investigations of Biber D., Johansson S., Leech G., in terms of English Grammar, everyday oral conversations are less lexically dense than those which refer to news reports [Biber D., Johansson S., Leech G. et al., 1999; 62]. While texts of news articles are carefully edited and proofed, 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**”. In average, in LSWE (Longman Corpus of Spoken and Written English), for example, lexical (meaningful) words comprise almost half of all words in this corpus. However, the proportion 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 D. et al., 1999; 65].

Such corpus statistics have become an indispensable part of grammar textbooks, dictionaries and other textbooks created for learners of a foreign language because they are significant in the process of development of professional competence of future language specialists. Furthermore, statistics visualization is also important for translation needs (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, statistic visualization serves as a reliable basis or source in researching the translation difficulties while comparing original texts and their translation [Tognini-Bonelly E., 2002; 67].

RESEARCH METHODS

Research questions

Taking into consideration all abovementioned corpus tools and technologies that may positively influence the English learning process, the following research questions arise.

1. Are there significant differences in the writing accuracy and complexity before and after the implementation of corpus-based assignments?
2. Which tools and technologies are easier to apply by students?
3. Which tools are more beneficial for learners to foster their writing development regarding the accuracy and complexity?
4. What are students' attitudes towards implementation of corpus tools and technologies in teaching writing?

Participants

The participants in this study were the 3rd year students of the 1st English language faculty, USWLU. Initially, there were 30 students, but 4 of them missed lessons several times and did not submit assignments on time, being, therefore, excluded from the experiment. Finally, in total, there were 26 participants, representatives of groups 1932 (experimental group) and 1936 (controlled group). According to a placement test, the level of participants was B1-B2 according to CEFR. The reason of taking this category of students was because they had two years of experience at university and were not novices in their institution.

The instruments such as largely closed-ended questionnaire with a limited number of open-ended questions, placement tests, set of corpus-based exercises and tasks and final tests were developed. The questionnaire was formulated on a 5-point

scale: *strongly agree; agree; neutral; disagree; strongly disagree.*

A placement test focused on differentiation of learners’ background knowledge of corpus technologies and their implementation in practice.

In total there are four stages of the experiment:

1. Introductory: explanation of the essence of corpus technologies in a language learning and discussion of the key points of an experiment organization;
2. Organizational and technical: registration of students in the COCA, demonstration of how some corpus technologies can be used;
3. Communicative: searching and studying examples given in communicative tasks to develop a writing competence of learners, discussing the identified examples, peer editing and making necessary changes in their written papers;
4. Reflective: evaluation of students’ assignments by the researcher and students’ written reflection on the experiment;

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).

Example of a concordance-based assignment

This task is aimed at mastering lexical units, which are difficult for common implementation and this task is recommended to students with B2 proficiency level. However, the concordance sorting tasks can be used from the earliest period of teaching; namely, it can be implemented not only for B2 level learners, but also for A2 and B1. Here, everything depends on the corpus 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 (Table 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
and “agriculture, “advocates a sort of soft resilience to withstand the forces of nature, not of brick or reinforced concrete but
Resilience: Learning From Food Security Crises # EU puts resilience at the heart of its work on fighting hunger and poverty # Today Commissioner
new policy to help vulnerable communities across the world build resilience to future crises. # The new Communication, “The EU Approach to
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

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 7).

Table 2. Combinability and frequency data of the phrase to 'build resilience'

1.	#to be raised when the economy is healthy to build resilience against losses and relaxed during a downturn to encourage greater lending. # Officials
2.	#they told themselves they could do it. To build resilience , praise effort and emphasize that the child simply has not mastered a skill
3.	#used to say the goal of CVE was to build resilience within the communities, but it was actually quite divisive within the community.
4.	#them where they are leaking time and how to build resilience to that. # First of all, you should look inside your company
5.	#to be. What I have to do is to build resilience into Barclays so that we can cope with those things, and where possible
6.	#their expectations, "Fredrickson says. You can build resilience with a steady diet of positive emotional experiences. That can be as simple
7.	#classrooms operated by the Allegheny Intermediate Unit help students build resilience is by following a daily routine. "Children like to know they are
8.	#and they do not necessarily help an individual to build resilience , to overcome difficulties as they happen or to react to challenges with composure
9.	#needs "with early intervention and work" to build resilience prior to combat deployments. "That is not true. "Navy medicine
10.	#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.

"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:

"In BBC article, in the simultaneous translation of a quote from a Russian politician, the verb **"exclude"** (*Russia's President Dmitry Medvedev does not exclude his participation in presidential elections for the next term*) is given in quotation marks

as not entirely appropriate to the context. In the rest of this BBC article, the preference is given to the verb **rule out**, despite the fact that the dictionary (Longman) defines these two verbs as synonymous [Gorina O., 2018; 191].

Task. Using the genre breakdown of the COCA, investigate the usage of these units 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.

Sociolinguistic variables’ analysis task

Sociolinguistic variables are accepted as differentiative tools when “on different occasions and for different purposes, English is used as a means of communication” [Gvishiani N., 2008; 53]. In such sort of 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 N., 2008; 53]. The expressions associated with a 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 (Tables 3, 4).

Table 3. Corpus data 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!
(SP:PS0FM)	It was taking a long time!
(SP:PS0FN)	What?
(SP:PS0FG)	They’re supposed to...
(SP:PS0FM)	Ha! Ha!
(SP:PS0FG)	I mean, to pay people, to come and clean the caravan out at the weekend.
(SP:PS0FN)	No, I mean this flue thing should’ve been...

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

Task. Students can be asked to compare formal/informal, academic/non-academic usage of lexical units paying attention to differences and similarities of the usage of vocabulary 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. Besides, 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 that of 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 (table 9). Besides such corpus texts give information about the time and place of the speech event, what the conversation is about, etc. The style and register features accompanying a given speech act, such as dialogue/monologue, informational/creative and others are also of great importance. These bits of knowledge present a complete discourse unit, which merges structural, semantic, and contextual dimensions [Tognini-Bonnelly E., 2002; 79].

Text analysis task

There are a lot of problems connected with a word usage which learners might face while learning a language. While being used in speech act, different word forms usually adhere to certain contexts in which they are used more or less regularly. In this case the frequency of usage of word-forms based on such features as their functional capacity, range of collocability, and utility can be differentiated. The most challenging here is that some forms of words may obtain a greater potential for implementation 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 *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 usage 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).

Task. Each student is assigned to find a pair of the most and the least frequently used words/phrases from the text under the analysis in COCA (Picture 1). 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.

Picture 1. COCA text analysis tool

The screenshot shows the COCA text analysis tool interface. The browser address bar displays www.english-corpora.org/coca/ha.asp. The interface has a navigation bar with tabs: SEARCH, FREQUENCY, CONTEXT, and ANALYZE TEXT. The ANALYZE TEXT tab is selected. Below the navigation bar, there are buttons for EDIT TEXT, SAVE TEXT, and radio buttons for WORD and PHRASE. A table shows frequency ranges and their percentages: 1-500 (55%), 501-3000 (14%), and > 3000 (19%). Below this, a text snippet is shown with the word 'tolerance' highlighted. To the right, a word sketch is displayed, showing associated words categorized by frequency: LOW FREQ, MID FREQ, and HIGH FREQ.

FREQ RANGE	1-500	501-3000	> 3000
369 WORDS	55 %	14 %	19 %

CLICK ON ANY WORD BELOW FOR A FULL WORD SKETCH

multiple etiologies, multiple controlling variables).

In both the undergraduate and graduate populations, low tolerance for ambiguity was associated with valuing course structure and with anxiety when the valued elements of structure were missing. These findings may be useful for instructors as they plan their courses. Instructors who are more flexible or unstructured in their approach to assigned readings or testing might want to construct profiles of class tolerance. In classes with high numbers of students with low ambiguity tolerance, instructors might reexamine the utility of employing a more structured approach. Future research is needed to identify the settings (e.g., specific career program fields) likely to have high rates of individuals with low tolerance for ambiguity.

(CLICK ANY WORD FOR FULL WORD SKETCH)

LOW FREQ	MID FREQ	HIGH FREQ
8: tolerance	3: associated, career	18: to
6: instructors	2: approach, classroom, graduate, multiple, structure	11: of, the
5: ambiguity	1: addressing, al, attitudes, benefits, client, content, despite, developing, e, elements, environment, et, eventually, findings, identify, impact, importance, likely, missing, necessary, organization, perform	10: in, with
3: anxiety	2: construct, explorative, flexible, undergraduate, unstructured	8: and
2: construct, explorative, flexible, undergraduate, unstructured	1: accomplishments, adapt, adapting, adjustments, affinity, ambiguities, ambiguity-tolerance, assigned, closely, coping, creativity, deadlines, desensitized, educators, emphasize, employing, etiologies, evaluations, everyday, exam, exploration, foster	6: a, that
1: accomplishments, adapt, adapting, adjustments, affinity, ambiguities, ambiguity-tolerance, assigned, closely, coping, creativity, deadlines, desensitized, educators, emphasize, employing, etiologies, evaluations, everyday, exam, exploration, foster		4: low, more, students
		3: are, course, for, have, levels, may, might, or, was
		2: as, be, both, fields, given, high, how, modeling, need, should, student, their, these, they, this, those, toward
		1: allow, also, an, at, become, been, better, by, class, classes, consider, controlling, courses

As it can be seen from the figure above, 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 N., 2008; 94].

RESULTS AND DISCUSSION

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. Yet, 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 the 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 enables both teachers and students to form skills of an autonomous researcher by having the access to a large, authentic reference system.

CONCLUSION

At the end, it is worth mentioning that the development of corpus competence of educators and learners through having an access to corpus-based data and corpus technologies as a source of authentic materials will improve and develop not only existing traditional methods of teaching foreign language, but also bring “nativeness” to the classroom. The reliance on the considerable potential of corpus technologies which have such characteristics of as authenticity, representation of different registers, linguistic and statistical visibility and a variety of search capabilities presented in a linguistic and methodological context, enables effectively develop all language competencies. Besides, corpus data “can aid in creating authentic situations” and “lessons involving authentic” data “can be featured to increase exposure to the language associated with each”. The activities highlighted in the research above, are the types of innovative approaches that can be incorporated in teaching materials. “Teachers will not have to locate and create materials and tasks, the ideas can be brought to them” [Radjabova G., 2021; 351]. In our research, we have provided an overview of the main directions of the development of the language competence. The experience

of compiling a small, professionally-oriented corpus based on instructional materials and presented in the article above suggests the cooperation of not only a team of teachers from one department, but also other disciplines experts. The feasibility of compiling a corpus of certain educational institution is oriented and aimed at selecting the most appropriate and exact materials to be studied.

REFERENCES

1. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics*, 6(1), 131-7.
2. Azimov, E.G., Shchukin, A.N. (2009). *New dictionary of methodological terms and concepts (theory and practice of language teaching)*. Moscow: IKAR.
3. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use*. Cambridge University Press.
4. Biber, D., Johansson, S., Leech, G. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.
5. 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*. Selected papers from the sixth international conference on teaching and language corpora. Amsterdam: Rodopi.
6. Francis, W.N. (1992). Language corpora B.C. In J.Svartvik (Eds.), *Directions in corpus linguistics*. Berlin: Mouton de Gruyter.
7. Gorina, O.G. (2013). Using corpus evidence to compile learner's vocabulary. *Bulletin of the Leningrad State University named after A.S. Pushkin*, 7(2), 201-209. (In Russian)
8. Gorina, O.G. (2014) Psychological grounds of using corpora in the ESP classroom. *Bulletin of the Leningrad State University named after A.S. Pushkin*, 7(1), 172-179. (In Russian)
9. Gorina, O.G. (2014). Using the technology of corpus linguistics for the development of lexical skills of regional students. *Cand. thes. in ped. scien*. Moscow.
10. Gorina, O.G. (2018). Tools of corpus analysis in teaching foreign language. *Tomsk State University Journal*, (435), 187-194.
11. Gvishiani, N.B. (2008). *Textbook on corpus linguistics: studies* (An English manual). Moscow: Higher School.
12. Gvishiani, N.B. (2010). *Introduction to contrastive lexicology (English-Russian interlanguage correspondence)*. Moscow: Librokom.
13. Johns, T. (1991). Should you be persuaded – two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal*, 4, 1-16.
14. 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). Saint Petersburg.
15. 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). London and New York: Longman.
16. Klepikova, T.A., Klushina, Z.V. (2012). The study of terminological collocations from the perspective of cognitive grammar and corpus linguistics. *TSU Bulletin*, 12, 284-288.
17. Mamontova, V.V. (2010). Corpus linguistics in modern paradigm. *Actual Issues of Modern Science*, (12), 230-238.
18. O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
19. Radjabova, G. (2021). The implementation of spoken corpora in creating teaching materials.

- International Journal on Integrated Education*, 4(5), 349-354.
20. Rakhmanov, I.V. (1972). *The main directions in the methodology of teaching foreign languages in the nineteenth and twentieth centuries*. Moscow: Pedagogy.
 21. Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
 22. Semich, Y.I. (2018). Didactic properties and methodological functions of corpus technologies. *Bulletin of Tambov University*, 23(177), 39-49.
 23. Shalamova, N.N., & Fil'chenko, A.Y. (2004). *Corpus linguistics and its use in the profile-oriented teaching of foreign languages*. Tomsk: TPU.
 24. Sinclair, J. (1991). Shared knowledge. In J. Alatis (Eds.), *Georgetown university round table on languages and linguistics*. Georgetown University Press.
 25. 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.
 26. Sysoev, P.V. (2010). Linguistic corpus in foreign language teaching methodology. *Language and Culture*, 1, 99-111.
 27. Ter-Minasova, S.G. (2009). *Speech syntagmatics: Ontology and heuristics: General and English syntagmatics of composite nominative groups* (2nd ed.). Moscow: Librokom.
 28. Tognini-Bonelli, E. (2002). *Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach*. Philadelphia: John Benjamins.
 29. Zakharov, V.P. (2011). *Corpus linguistics: textbook for students of humanitarian universities*. Irkutsk: IGLU.