

UDC 378.147

## THE ROLE OF CORPORA IN FOREIGN LANGUAGE EDUCATION

*Mikhaylova A. G.*

*Sevastopol State University, Russian Federation*  
<https://orcid.org/0000-0002-9419-543X>

Currently the problem of corpus approach application as digital means to solve foreign language education issues are relevant due to foreign language study enhances one's opportunities in technology, military, business, government, medicine, law, industry, marketing, etc. The purpose of the study: to identify the main role of corpora in teaching a foreign language. The following research methods are: content analysis of scientific literature on the research topic, interpretation, systematization, structuring. Language education as a process of cognitive activity aimed at mastering language and speech, at self-development and personality formation. It is based on the ability and readiness formation to participate in oral, written, virtual, intercultural, organizational, corporate, scientific communication using foreign languages as a means. In the context of foreign language education high requirements for teaching practical skills of translating texts in the specialty and special terminology are put in. It is necessary to create a connection between the digital society and the humanities education, which is supposed to be a digital culture. There are technologies associated with the routine process of developing work programs, means, etc. Expanding the corpus capacity to translate texts contributes to the diversification of learning processes through digital technologies.

**Keywords:** digital society, corpus, translation, training, foreign language.

### PROBLEM STATEMENT

The concept of humanization of higher education is originally understood as protection against the technization of society, not accidentally called the scientific and technological revolution.

Foreign languages translation is founded on an interdisciplinary basis [6]. It is aimed at “development of communicative, cognitive, sociocultural, informational, professional and general cultural competencies...” [9, p. 147]. The ability to translate professional texts is valuable for specialists [15]. The task of educators is to teach students to translate texts taking into account the factor of cultural differences.

Today, “the concept of correspondence and/or equivalence has been a complex issue among translation experts and theorists due to the cultural dimension involved in the languages involved (source language and target language)” [2, p. 50].

More essentially, teaching corpus features to university students would be useful as it could affect the quality of translated text. Nevertheless the genre frequency of collocation analysis (for example, “adj. + problem”) would be an important aspect in foreign language education that should be emphasized in academic classes.

The paper aims to investigate if compiling a corpus of personal statements can lead to creating an effective corpus-based activities to be taught in teaching writing a personal statement. T

Corpus linguistics was studied by different authors [7, 10, 11, 16, 25]. The practice issues of corpus linguistics were stated by D. Biber, S. Conrad, R. Reppen (1998) [4], M. Baker (1999) [3], Kyung Hye Kim [14], S. Johansson [12], M. McCarthy [18], S. Voelkel, F. Kretzschmar [25] and others.

Analysis of recent achievements and publications shows that translation technology is the software tools use to facilitate text conversion from one language to another which can increase efficiency, productivity in managing multilingual text content.

The problem in general can be defined: what are the effective digital means of text interpretation? Its connection with important scientific and practical issues is in the fact that digital educational technologies are an innovative way of the educational process realizing [13]. It is based on the electronic systems application that provides visibility [17].

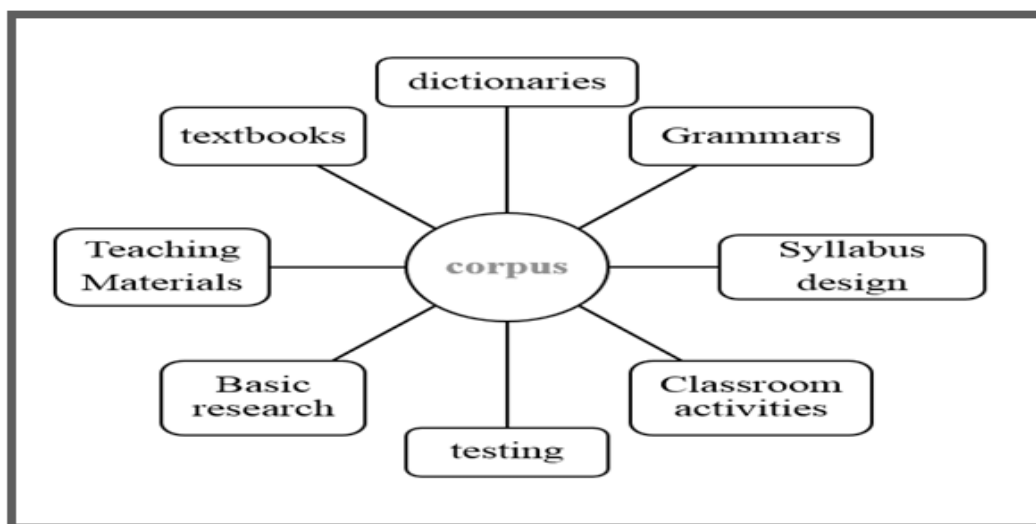
The novelty of the study is in solution problems of translation and interpreting studies using the corpus approach. The paper analyzed a part of the concordancer for the verb crack and the frequency of collocation “adj. + problem” in the Second or Successive Habeas Corpus Petition (SOS) case corpus. The results can be used to create corpus-based excersises.

**The purpose.** In this paper we study the issues of corpus approach using to solve translation problems. Language is to be considered as an essential part of understanding of the world. It affects science, culture and religion. The barriers occured have to be diminished by means of translations tools. The following tasks are to be solved: to define the role of electronic corpora in appropriate selection of vocabulary; to find a type of language corpora to process text information. The Corpus of Contemporary American English (COCA) was chosen as the corpus program for this study.

#### THE MAIN RESEARCH MATERIAL

K. Bouziane wrote about cultural transfer in advertising. She proposes the strategies in the transferring of verbal elements non-verbal elements in advertising from using semiotic tools [5]. A. Negoescu writes about discourse aspects. When considering language in use in real situations these boundaries between the levels of the language are blurry [23].

Many scientists are “strongly arguing for the effective relevance of corpora in language teaching” [1, p. 108]. Corpora could impact language teaching in the various areas (picture 1).

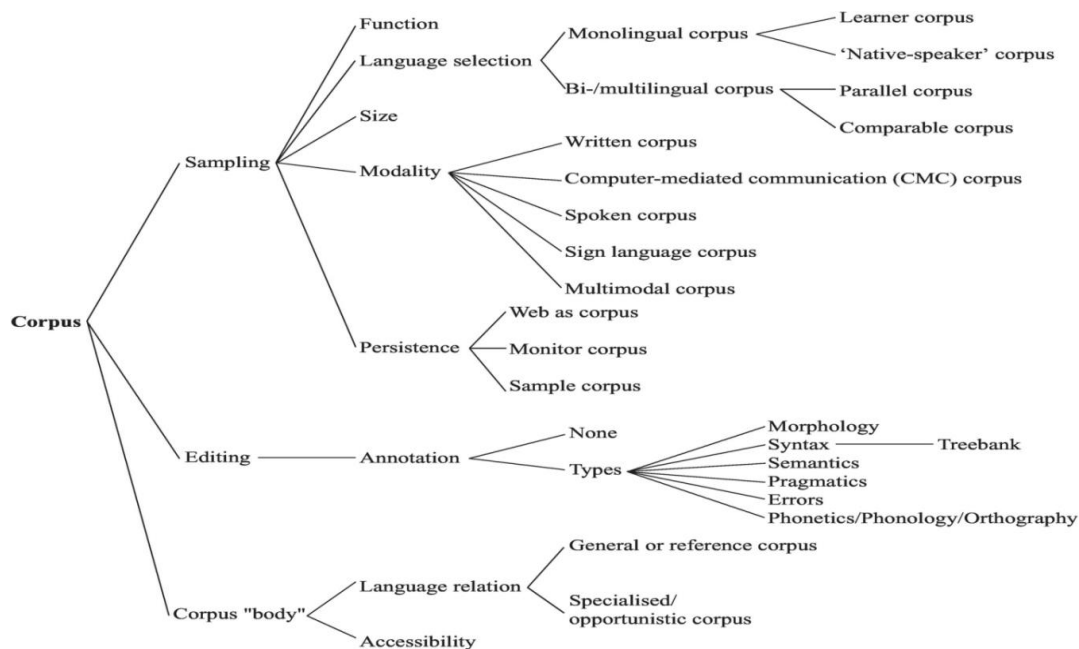


Picture 1. Corpora influence on language teaching [25].

A. Saricoban believes that interpreting requires realization of native language and target one in “both language structures and cultures exploring similarities and differences between them” [24, p.4]. “The adequacy of the translation is exhaustive accuracy in the transfer of the semantic content of the original and a full-fledged functional and stylistic correspondence to it” [20, p.88].

Interpreters should identify factors which can affect the quality of the text. “The main role of corpora is to provide the most correct examples of the use of language units...” [21, p. 361]. A corpus as a set of texts processed is applied as a basis for the study of a language (confirmation of linguistic rules in a given language; testing of statistical hypotheses; statistical analysis). We perceive the world through our senses, memories, the culture to which we belong and the language we happen to speak. So, language is considered to be a major part of our understanding of the world as it provides the interaction between persons [4]. “Culture, religion and science depend on language” [22, p. 366] and language carries them all. Therefore language is necessary premise to differentiate worlds and the barriers among languages are to be lessened through translations or interpretation.

A corpora-based approach is used to cluster short texts, since they give rise to the problem of insufficient word matching and lack of contextual information (picture 2). The feature of the corpus is a collection of texts available for translated material analysis [8, 14, 22].



Picture 2. Types of corpora linguistics (by S. Voelkel, F. Kretzschmar) [25].

THE ROLE OF CORPORA IN FOREIGN LANGUAGE EDUCATION

“Corpus linguistic approaches are more integrated with other methodologies used in applied linguistics. There are many ways to define a corpus. Scholars will define a corpus as a collection of machine-readable authentic texts that are selected as being representative of a particular language or variety of language” [19, p.12]. Thus, “a corpus is a set of texts, written or oral, that are stored on a computer” [3, p. 281].

Processing allows to search for any necessary information in the case. In corpora, as a rule, the markup of texts is carried out by parts of speech, this allows you to find words in specified grammatical forms and in specified structures and combinations (collocates). Special concordance programs are used to process information. They search for information in an array of texts and form a concordance — a list of all contexts in which a word or phrase occurs in the text under study. Here, for example, is a part of the concordancer for the verb crack (color markers show parts of speech in the immediate environment of the searched word) (see picture 3).

WEBSITE		SORT	SORT	SORT
151	lmsdb.com	the hell it 's made of , but <b>ca</b> n't	<b>crack</b>	<b>is</b> <b>used</b> DONALD Good . Clear it off the table so
152	bleubirdblog.com	and cheese . Pick out your favorite bottle of <b>wine</b> and	<b>crack</b>	<b>is</b> <b>used</b> Light a nice- candle- to set the mood .
153	lwantoneofthose.com	they deserve every last drop if they do <b>finally</b> <b>manage</b> to	<b>crack</b>	<b>is</b> <b>used</b> they do n't , aside from you crowing ,
154	hpc-uk.org	was said and got frustrated and kicked the <b>car</b> <b>window</b> ,	<b>cracking</b>	<b>is</b> <b>used</b> again Mother A felt the situation was made worse
155	eoshd.com	do ! With all of Adobes resources youd <b>think</b> <b>theyd</b> <b>have</b>	<b>cracked</b>	<b>is</b> <b>used</b> # Apples decision to- start afresh with FCPX on the
156	www.all.com.au	people in that position probably would have had the <b>right</b> to	<b>crack</b>	<b>is</b> <b>used</b> # " I thought , you know what ? There
157	futureofpersonalhealth.com	for good must be our next accomplishment . <b>if</b> <b>we</b> <b>can</b>	<b>crack</b>	<b>is</b> <b>used</b> <b>its</b> <b>molecular</b> <b>code</b> , we will have extraordinarily scalable
158	sciencemuseum.org.uk	of hydrogen and oxygen , marking the <b>first</b> <b>step</b> towards	<b>cracking</b>	<b>is</b> <b>used</b> <b>its</b> <b>molecular</b> <b>structure</b> . Each molecule consists of one oxygen
159	venturebeat.com	to be somewhat recognizably human . Characters <b>can</b> <b>swear</b> and	<b>crack</b>	<b>jokes</b> <b>and</b> <b>possess</b> active sex drives . It 's not like ,
160	techcrunch.com	around the world . Have them play DJ while <b>you</b> <b>can</b>	<b>crack</b>	<b>jokes</b> <b>between</b> <b>songs</b> . # 's most popular station Radio Mozart
161	oscarfishlover.com	like he is protecting something other than territory . <b>i</b> <b>always</b>	<b>crack</b>	<b>jokes</b> <b>that</b> <b>he</b> <b>is</b> " playing with himself " on the rocks
162	miaminewtimes.com	seamlessly to working as a writer . He <b>hustled</b> . He	<b>cracked</b>	<b>jokes</b> <b>he</b> loved to talk . And talk . And talk
163	pcgscolinfacts.com	# Die State III : As above , <b>but</b> <b>with</b> <b>die</b>	<b>crack</b>	<b>just</b> <b>beginning</b> <b>at</b> star 12 ; hairline crack extends to border ;
164	rouxbe.com	made these a few times now and the <b>topps</b> <b>are</b> <b>always</b>	<b>cracked</b>	<b>kind</b> <b>of</b> <b>look</b> like an oatmeal raisin cookie . What should I
165	nathanbransford.com	up . The Coldwater River . Confirming her <b>fears</b> , <b>ice</b>	<b>cracked</b>	<b>loud</b> <b>as</b> <b>a</b> pistol shot . Carter undid his seatbelt . Tori
166	grc.com	on this exponential rise , 2021 was where the <b>cost</b> to	<b>crack</b>	<b>made</b> <b>it</b> <b>begin</b> to be feasible for nongovernmental agencies .
167	puzzlepirates.com	is fully intact (those who lobby for <b>overhauled</b> <b>graphics</b> <b>also</b>	<b>crack</b>	<b>me</b> <b>up</b> <b>and</b> , as they completely miss the point that the
168	mixtape torrent.com	neva trapped a day in his life ! ! <b>Yaw</b>	<b>crack</b>	<b>me</b> <b>up</b> <b>cause</b> yaw 2 Lame ! ! Blow pops . / .

**Picture 3.** Part of the Concordancer for the Verb Crack  
Source: <https://www.nottingham.ac.uk/alzsh3/acvocab/concordances.htm>

The above example is from the Corpus of Contemporary American English (COCA) (picture 4).

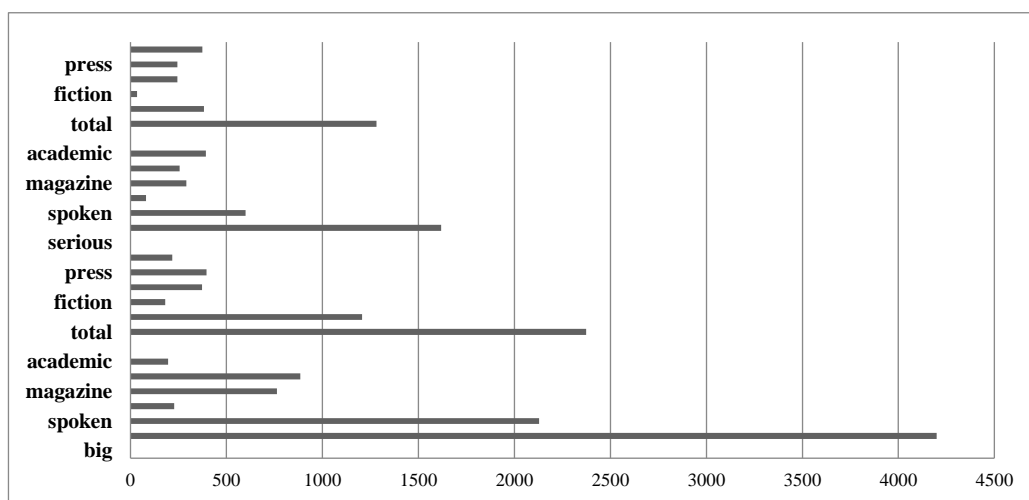
Different corpora contain various types of texts (written only, oral only, both types) and different sets of genres. COCA presents a wide range of genres.

Genre	# texts	# words	Explanation
Spoken	44,803	127,396,932	Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Oprah)
Fiction	25,992	119,505,305	Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books 1990-present, and fan fiction.
Magazines	86,292	127,352,030	Nearly 100 different magazines, with a good mix between specific domains like news, health, home and gardening, women, financial, religion, sports, etc.
Newspapers	90,243	122,958,016	Newspapers from across the US, including: USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. Good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.
Academic	26,137	120,988,361	More than 200 different peer-reviewed journals. These cover the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business
Web (Genl)	88,989	129,899,427	Classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages (by Serge Sharoff). Taken from the US portion of the GloWbE corpus.
Web (Blog)	98,748	125,496,216	Texts that were classified by Google as being blogs. Further classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages. Taken from the US portion of the GloWbE corpus.
TV/Movies	23,975	129,293,467	Subtitles from OpenSubtitles.org, and later the TV and Movies corpora. Studies have shown that the language from these shows and movies is even more colloquial / core than the data in actual "spoken corpora".
	<b>485,179</b>	<b>1,002,889,754</b>	

**Picture 4.** Summary of Data on COCA  
 Source: <https://www.english-corpora.org/coca/>

COCA is the balanced and large corpus of American English related to many other corpora of English. It is probably the most widely-applied corpus of English. Formerly these corpora were known as the “BYU Corpora”. They offer unparalleled comprehension into variation in English.

The frequency of collocation “adj. + problem” in the SOS case corpus is shown in the figure 1 (for adj. big, real, serious, major).



**Figure 1.** The collocation “adj. + problem” frequency in the SOS case corpus

Corpus instruments have enabled linguistic educators to study actual usages and the features of certain genres to enhance syllabus design and infer more effective exercises. For genres that are highly influential, corpus tools are useful.

Further analysis showed that the most frequent adjective are 'big' and 'real'. Interestingly, all four adjectives (big, real, serious, major) have different variations of translation in genres.

### CONCLUSION

In the context of foreign language education high requirements for teaching practical skills of translating texts of different genres and specific terminology are put in. A new corpora-based approach is used to cluster short texts, since they give rise to the problem of insufficient word matching and lack of contextual information. Corpora impacts language teaching in different fields for its possibility to present proven evidence of the used language statistically.

Corpus-based approach to foreign language teaching has demonstrated its benefits in different research. Corpus can present various effective ways of text interpreting. Data integrated with lines of concordances allow opportunities for linguistic research.

Corpus means have contributed in detection of the various lexical features behavior.

### Список литературы – References

1. *Almutairi N. D.* The Effectiveness of Corpus-Based Approach to Language Description in Creating Corpus-Based Exercises to Teach Writing Personal Statements // *English Language Teaching*. – 2016. – No. (7). Vol. 9. – P. 103.
2. *Anthony N., Zayyanu Z. M.* Achieving Correspondence and/or Equivalence in Translation, An Evaluation of the Translation Ekwensi's Burning Grass into French as La Brousseardente by Françoise Balogun // *Journal of Research in Humanities and Social Science*. – 2021. – Vol. 9, Issue. – P. 50–55.
3. *Baker M.* The role of corpora in investigating the linguistic behavior of professional translators // *International Journal of Corpus Linguistics*. – 1999. – No. 4. – P. 281–298.
4. *Biber D., Conrad S., Reppen R.* *Corpus linguistics: Investigating language structure and use*. – Cambridge: Cambridge University Press, 1998.
5. *Bouziane K.* Cultural Transfer in Advertising: A Semiotic Approach in Investigating Translated Advertisements from English to Arabic: The Case of Cosmetics // *Professional communication and translation studies*. International conference. – 7 edition. – 2011. – Available from: <https://docplayer.org/60395657-Professional-communication-and-translation-studies.html>. – (Accessed 21 February 2024).
6. *Chroma M.* The role of translation in professional communication // *The Routledge handbook of language and professional communication*. – 2014. – P. 147–164
7. *Chul-Kyu Kim.* Personal pronouns in English and Korean texts: A corpusbased study in terms of textual interaction // *Journal of Pragmatics*. – 2009. – V. 41, Issue 10. – P. 2086–2099.

8. Chu Tao Zheng, Cheng Liu, Hau San Wong. Corpus-based topic diffusion for short text clustering *Neurocomputing*, 2018. V. 27531. Pp. 2444–2458 <https://doi.org/10.1016/j.neucom.2017.11.019>
9. Foreign language for non-linguistic universities and faculties. Sample program / Edit. by S.G. Ter-Minasova. – Moscow, 2009. – 24 p.
10. *Gilmore A., Millar N.* The language of civil engineering research articles: A corpusbased approach // *English for Specific Purposes*. – 2018. – V. 51. – P. 1–17.
11. *Hajiyeva K.* A corpus-based lexical analysis of subject-specific university textbooks for English majors. – *Ampersand*, 2015. – V. 2. – P. 136–144
12. *Johansson S.* On the role of corpora in cross-linguistic research // *Corpora and crosslinguistic research*. – Amsterdam, 1999. – P. 3–24.
13. *Kuznetsova T.F.* Digitalization and digital culture // *Horizons of humanitarian knowledge*. – 2019. – No. 2. – P. 96–98
14. *Kyung Hye Kim* Newsworld discourses on China and their Korean translations: A corpus-based approach. // *Discourse, Context & Media*. – 2017. – Vol. 15. – P. 34–44.
15. *Lekareva V.* Historical facts about London bridges // *Recent Achievements and Prospects of Innovations and Technologies*. – 2019. – P. 457–464.
16. *Li X.* Identifying in-demand qualifications and competences for translation curriculum renewal: a content analysis of translation job ads // *The Interpreter and Translator Trainer*. – 2022. – Vol. 16. Issue 2. V. 16. – Available from: <https://www.tandfonline.com/doi/full/10.1080/1750399X.2021.2017706> – (Accessed 21 February 2024).
17. *Macroeconomics: textbook – method. Allowance.* Minsk: Ed. center of BSU, 2012. 159 p.
18. *McCarthy M.* Accessing and interpreting corpus information in the teacher education context // *Language Teaching*. S.I. Vol. 41. – Cambridge: Cambridge University Press, 2008. – P. 563–574.
19. *McEnery T., Xiao R., Tono Y.* *Corpus-based language studies: An advanced resource book*. – London: Routledge, 2006. – 91 p.
20. *Mikhaylova A.G.* The development of students' professional creative skills in foreign language learning // *Язык: категории G, функции, речевое действие. материалы девятой научной конференции с международным участием: в 3 частях*. – 2016. – С. 87–91.
21. *Mikhaylova A.G.* Corpus approach to theoretical and applied aspects solution of military interpretation // *Recent Achievements and Prospects of Innovations and Technologies*. – 2020. – P. 360–365.
22. *Mikhaylova A.* Corpus approach to theoretical and applied aspects solution of military interpretation // *Recent Achievements and Prospects of Innovations and Technologies*. – 2020. – P. 365–369.
23. *Negoescu A.* Aspects of advertising discourse // *Professional communication and translation studies. International conference*. – 7 edition. – 2011. – Available from: <https://docplayer.org/60395657-Professional-communication-and-translation-studies.html>. \_\_\_\_ – (Accessed 21 February 2024).
24. *Saricoban A.* Translation to Test Language Areas and Skills // *Professional communication and translation studies. International conference*. – 7 edition. – 2011. – Available from: <https://docplayer.org/60395657-Professional-communication-and-translation-studies.html>. \_\_\_\_ – (Accessed 21 February 2024).
25. *Voelkel S., Kretzschmar F.* *Corpus Linguistics*. In *Introducing Linguistic Research*. – Cambridge: Cambridge University Press, 2021. – P. 133–165.

## ОСНОВНАЯ РОЛЬ КОРПУСОВ В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

*Михайлова А. Г.*

Гуманизация высшего образования интерпретируется как начало формирования обычного общества и цифровой культуры. Её идеи, высказанные еще до реализации в цифровой экономике, изначально понимались как некая защита от технизации общества (образования в частности). В настоящее время проблема применения корпусного подхода как цифрового средства решения задач обучения иностранному языку актуальна в связи с тем, что изучение иностранного языка расширяет возможности человека в технике, армии, бизнесе, правительстве, медицине, юриспруденции, промышленности, маркетинге и т. д. Цель исследования: выявить основную роль корпусов в обучении иностранному языку. Методы исследования: анализ научных источников по теме исследования, интерпретация, систематизация, структурирование. Языковое образование как процесс познавательной деятельности направлен на саморазвитие и формирование личности. В его основе – формирование умения и готовности к участию в устном, письменном, межкультурном, виртуальном, организационном, корпоративном, научном общении с использованием в качестве средства иностранных языков. В условиях иноязычного образования предъявляются к обучению практическим навыкам перевода текстов по специальности и специальной терминологии высокие требования. Необходимо создать связь между цифровым обществом и гуманитарным образованием, которым должна быть цифровая культура. Существуют технологии, которые связаны со сложным процессом разработки рабочих программ по дисциплинам, средств, но есть и понимание культурологической деятельности по гуманизации высшего образования. Расширение потенциала корпуса по переводу текстов способствует диверсификации возможностей процессов обучения за счет цифровых технологий.

**Ключевые слова:** цифровое общество, корпус, перевод, обучение, иностранный язык.