РОЗДІЛ 2

ФОНЕТИЧНА, ЛЕКСИЧНА ТА ГРАМАТИЧНА СИСТЕМИ МОВИ ТА МЕТОДИ ДОСЛІДЖЕНЬ

Отримано: 13 листопада 2018 р.

Прорецензовано: 20 грудня 2018 р.

Прийнято до друку: 21 грудня 2018 р.

e-mail: anabel003@i.ua

DOI: 10.25264/2519-2558-2018-4(72)-69-73

Belyaeva Anastasiia. Corpora creation in contrastive linguistics. *Наукові записки Національного університету «Острозька академія»: серія «Філологія»*. Острог : Вид-во НаУОА, 2018. Вип. 4(72), грудень. С. 69–73.

УДК 811.111:811.161.2.091:81'42

Anastasiia Belyaeva,
PhD, Associate Professor, Zaporizhzhya National University

CORPORA CREATION IN CONTRASTIVE LINGUISTICS

Universal and specific features of language usage can become more evident if tested against the non-elicited language data on large scale. This requirement can be met by using corpora that provide ample data to test research hypotheses in contrastive language studies in objective and falsifiable manner. However, criteria in corpora creation and comparability measures in the evaluation of available corpora present a separate problem in contrastive linguistics. The article presents an overview of the types of corpora used in Contrastive Linguistics research and describes their characteristic features. The study proceeds to look into the sources of data used in corpora creation both in (commercially) available corpora and data collections compiled to answer a particular research question. The article describes the techniques used in creating comparable corpora for contrastive studies and presents the comparability measures to evaluate the corpora. The study examines the case of building a topic-specific comparable corpus in English and Ukrainian. The corpus focuses on education-related vocabulary in the languages under analysis. The corpus comparability is measured using translation equivalence and word frequency similarity. The article used the procedures outlined above to collect a quasi-comparable (non-aligned) corpus focusing on the topic of education with the English and Ukrainian languages in contrast. Using frequency comparability measure it was established that both components of the corpus (in the English and Ukrainian languages) contain keywords related to the topic of education.

***Key words**: monolingual corpus, parallel corpus, contrastive linguistics, comparability measure.*

Бєляєва Анастасія Вікторівна,
кандидат філологічних наук, доцент, Запорізький національний університет

СТВОРЕННЯ КОРПУСІВ У ДОСЛІДЖЕННЯХ З ЗІСТАВНОГО МОВОЗНАВСТВА

У статті проаналізовано типи корпусів, які використовуються у дослідженнях з зіставного мовознавства з метою виявлення універсальних та специфічних особливостей мов. Встановлено основні джерела матеріалів для укладання корпусів, критерії відбору текстів, етапи укладання корпусів, моделі оцінки та характеристики корпусів для контрастивних студій. У статті розглянуто методи, що використовуються у створенні корпусів для зіставних досліджень, описано досвід укладання корпусів для зіставних досліджень на матеріалі англійської та української мов. Критерії відбору матеріалу, етапи побудови корпусів та перспектив їх використання розглянуто на прикладі корпусів лексики сфери освіти в аналізованих мовах.

***Ключові слова**: одномовнй корпус, паралельний корпус, контрастивна лінгвістика.*

Studies carried out on large language samples can give new insights into language usage. Corpora, collections of written text or speech, provide ample data to test research hypotheses in objective and falsifiable manner. Researchers interested in contrastive language studies take a corpus-based approach since universal and specific features of language usage can become more evident if tested against the examples retrieved from corpora. One approach towards incorporating corpora data to contrastive studies would be to use monolingual corpora to retrieve data on language usage and then look into similarities and differences in patterns evident in languages compared. However, in this case researchers could not be sure whether the monolingual corpora are based on equal criteria of collecting language samples. Therefore, the presence or absence of a certain language phenomenon could result from the type of documents that make a corpus. This article describes approaches to corpus creation and evaluation in contrastive linguistics research. The objective of the article is studying the techniques of corpus collection and analysis in contrastive language studies. The paper begins with an overview of approaches to materials selection for corpus-based contrastive research. The study proceeds to determine how corpora can be evaluated in order to be used in contrastive linguistics research. Finally, the paper attempts to examine the procedures of creating a corpus for research on the vocabulary in the field of education with English and Ukrainian languages in contrast.

Contrastive studies can rely on two types of multilingual corpora: parallel and comparable corpora. Scholars stress that in contrastive research multilingual corpora are of particular importance in studying language meaning [10]. The first type is represented by the original texts and their translations into one or several languages with sentences aligned. Such corpora can be used in translation studies or cross-language information retrieval. Scholars also single out noisy parallel corpora, which contain aligned sentences of original texts and their rough translations. The translations cover the same topic as source texts but contain alterations to the original, i.e. some passages or sentences might be added or omitted [4].

Comparable corpora, in turn, consist of original texts in two or more languages. In comparable corpora documents cover similar topics or belong to the same time period (e.g. newspaper articles published on the same date or in the same year). Such corpora should be aligned; however, criteria for the alignment present a separate research problem. Most often common topic is used as the basis for alignment. Another type of comparable corpora contains non-translated documents covering the same or different topics. Such corpora are heterogeneous collections of documents that do not require topic alignment [4]. It is the latter type of corpora that might be particularly useful for contrastive studies because it allows looking into non-elicited language data on large scale to make conclusions about language patterns.

Parallel corpora are widely exploited in translation studies; however, their use in contrastive research might prove disadvantageous. The primary concern is that such corpora tend to have smaller language coverage than monolingual or comparable corpora. What is more, few parallel corpora are readily available for researchers; their creation is time-consuming, unnecessarily laborious, and more often than not involving the subjective factor of human translator interference [7; 13].

Therefore, more linguists turn to comparable corpora and select the material for their research in several languages but with similar topic following similar principles. Yet, comparable corpora present a problem for linguists because there are no unified criteria for their creation and evaluation. The general requirement for corpora creation applies to comparable corpora as well: they should be representative and balanced [10]. It means that a good corpus should include all types and genres of text or speech that can help answer a particular research question, and the number of the text or speech types should be roughly equal for each language under study. What is more, in case of comparable corpora scholars should follow similar sampling techniques for each language under study [10, p. 134]. However, those are general guidelines, and the approaches individual researchers take to corpus collection largely depend on the question they seek to answer as well as on the type of data in different languages available or suitable for corpus creation.

Unless researchers are interested in learner language, materials for comparable corpora can be retrieved from the web. Comparable corpora are created from web materials using either automated queries or ‘web spiders’, software built for linguistic queries of the Internet. The first technique relies on either search engines (such as Yahoo, Google, etc.) or social media (such as Twitter) service called APIs. The service allows researchers to collect web pages or documents (however, there is a limitation on a number of pages downloaded per day) that contain a pre-defined set of words. The set of query words, in turn, can be compiled by the researchers using frequency dictionaries if they aim to create a corpus of general language, or it can be restricted to query words related to a specific topic (e.g. ‘Arab Spring’ in the study by Hajjem et al.) [7]. In this approach a corpus for each of the languages under analysis is built separately. Linguists begin by creating a set of query terms in one of the languages and then use it to compile a corpus in this language. If the task is to create a general language corpus, i.e. the researchers are not interested in a particular topic, the same procedure is repeated for other language(s) under study. Comparable corpora are also built from the search engines metadata serving as a classification system for the linguists. In that case researchers rely on information provided by webpage codes to retrieve query terms and find their possible equivalents in other languages. However, if linguists need a topic-specific set of data they consult bilingual dictionaries to compile a set of query words for each of the languages under analysis. Since the manual proceeding of web queries to build a corpus can be time-consuming, scholars automate this process by creating software that automatically retrieves the web pages containing the set of query terms [1]. In this approach corpora in two or more languages under analysis are compiled sequentially, not simultaneously.

Among the methods used to construct corpora from web documents are cross-language information retrieval and clustering. The first method relies on the use of keywords in the source language that are later translated into the target language and run against the target language document collection [13]. In this approach comparable corpora are treated as collections of texts on similar topics. The texts are retrieved from dissimilar sources, but they share a number of terms that are translations of each other [13]. The analysis begins with retrieving keywords from the source document using the RATF (relative average term frequency) formula, i.e. calculating the number of times a term occurs in every document in the text collection compared to the number of documents in the corpus. Then these key words can be translated into target language using bilingual dictionaries, dictionary-based translation programs or online translation systems [7; 13].

Newspaper articles and news agency reports are often used to create comparable corpora, since this type of documents can meet the criteria of similarity in composition, genre, topic, and communicative function. Scholars select either a particular topic or specific publication type in two or more languages and collect monolingual text data in each of the languages separately. In some cases newspaper articles retrieval can be combined with the techniques of search engines metadata analysis if the web pages (e.g. Zientzia.net that gives access to scientific publications in different languages) contain meta links to similar articles written in several different languages [13].

Wikipedia is another source widely used in comparative studies and comparable corpus building. Articles in two or more different languages can be selected through web crawling (using software developed by researchers themselves) from a pre-defined restricted domain (set of articles covering a specific topic) and further automatically aligned according to the similarity of the topic covered. Such an approach relies on machine translation and sentence-alignment programming tools. An alternative technique uses interlanguage links or tags provided in the code of Wikipedia articles. In this approach the corpus obtained can cover a wide variety of topics [5; 11].

In order to be used in contrastive linguistics research corpora should be comparable. There is no unified approach towards the definition of comparability and measures to establish it depend on the researcher’s tasks [9]. There are two groups of criteria used: qualitative or quantitative comparability criteria. Qualitative criteria take into account stylistic features of the documents that make up the corpus, namely the time the documents were created, their topic, genre, media they were created for and appeared in. However, this measure is most relevant at the stage of corpus collection because homogeneity and representativeness are among the basic criteria for both monolingual and multilingual corpora [10]. Stylistic features are also important if the task of the researcher is to create an aligned comparable corpus, in this case topics and dates could provide common ground for organizing documents in different languages.

Quantitative criteria rely on Cross-Language Information Retrieval (CLIR) comparability statistical measures and take into account the frequency of specific linguistic phenomena. Such measures involve analysing the quantity of common vocabulary in corpora in two or more languages and are often used in machine translation research. Linguists create a list of meaningful or frequent language units in a corpus in one of the languages under analysis. Techniques involved include lemmatisation and POS-tagging of each document in a text collection in one of the languages, which would be treated as a source corpus and provide a list of source words. Alternatively, comparability measures can make use of binary and vector models (cosine similarity) [7]. The first method uses a formula to calculate absence or presence of keyword translations in text collections in each of the languages under analysis; both text collections are seen as a bag of words and are treated as source and target corpora respectively. The second measure represents each document as a vector. The key indexes for text collections are translated from one language under study into another using machine translation, and vectors representing the weight of words in source and target documents are compared [7]. Depending on the size of the corpus, it can include all documents or a sample of randomly selected documents. Then linguists consult bilingual dictionaries or translation programs to establish the translations of the key vocabulary in question [13]. The two corpora are considered comparable if they meet the minimal criteria for the expectation of finding translations of source words in a target corpus [7; 12]. This technique is referred to as 'translation comparability measure'. The method relies on bilingual dictionary or translation system and may depend on the dictionary coverage or subjective factor in case of translation ambiguity problem. An alternative approach takes into consideration not only presence or absence of translations in corpora documents, but also thematic reference of the vocabulary under analysis [8]. The technique takes into account the number of occurrences of lexical units and their respective translations.

Quantitative measures can rely on comparison of the contexts words are used in or on the key words themselves. In the first case comparability measure seeks to establish similarity between documents in corpora looking into the words used in the same contexts. In this view, every document in the corpus is statistically analyzed to retrieve the most representative words it contains. The latter becomes the foundation for the comparison across languages. The similarity measure of usage context is complemented by the comparison between patterns lexical units appear in [4].

Another approach focuses on word frequency lists. In this approach the keywords of each corpus are established and statistics is used to measure similarity between the corpora. Statistical measures can rely on Chi-square (x^2) or log-likelihood tests [3; 9]. The former allows studying differences between actual occurrence of the words and their expected frequencies. Log-likelihood can be used to measure significant difference in frequency between two corpora and is regarded among the most widely used and most reliable statistical measures in keyword extraction [6]. In this view the rate of the most frequent words usage or word co-occurrence can be regarded as the quantitative measure of assessing similarity between corpora. Word frequency is of particular importance if researchers want to look into the key or most important /specific units in corpora of the languages in contrast. The concept of keyness becomes central if the study aims at analyzing the way a certain real world phenomenon or abstract notion is viewed by the speakers.

The present study focuses on verbalized representations of views that speakers of English and Ukrainian languages have about the phenomenon of education. To do it the study retrieves and compares key lexical units in English and Ukrainian language subcorpora in quasi-comparable corpus covering the topic of education. The research corpus contains texts that either belong to the sphere of education or discuss this phenomenon. The study uses a technique of building a comparable non-aligned corpus from the Web. The research attempts to make the corpus representative by controlling the pages and domains used as the source of data and includes different genres of written text. Documents for the corpus are collected using query terms or seed words which are randomly combined. The query terms are run through BootCaT web-based toolkit available via SketchEngine following the procedure outlined in Baroni and Bernardini [1]. However, unlike the cited study, which aimed at creating a corpus of General English, the present research does not use query terms randomly obtained from the search engines. Instead, the set of seed words for a corpus is related to its subject matter that is the concept of education. The concept of education verbalizes the phenomenon of the real world which possesses a complex structure. In present research a single seed word cannot be used to create a topic-specific comparable corpus, unlike the approach taken by Hajjem et al. [7] focusing on a single query term, 'Arab Spring'. The present study relies on a number of lexical units that do not directly name the concept of education, but are logically relevant to naming the phenomenon under analysis. Therefore education can be seen as a system of concepts that exists as a combination of interrelated parts. The concept of education is logically connected to conceptual fields consisting of typologically and semantically hierarchically ordered verbalized concepts in the languages under analysis. Such systems reflect the organization of the corresponding cognitive semantic space. The conceptual system under analysis is a complex multidimensional structure; it has its own hierarchy which reflects relations between real world phenomena.

In the present research the corpus for the study of lexical representation of concept EDUCATION in English and Ukrainian languages is built on the basis of seed lexemes. Since the purpose of building the corpus is to collect documents that reflect ideas the language communities have about the phenomenon of education, seed words used for compiling the corpus were not the most frequent words in the languages under analysis, but the lexical units reflecting the concept of education in the English and Ukrainian languages. The lexical units belonging to the conceptual field under analysis were retrieved from English and Ukrainian monolingual dictionaries. Thematic classification of lexical units that belong to the conceptual field of education resulted in singling out the following groups: I) educational establishments: *school* [17, p. 1273–1274]; *школа* [14, p. 1005]. In this class there are lexical units that name educational establishments according to: a) contents of education: *technical college* [17, p.284]; b) organization of educational process: *boarding school* [17, p.153]; c) type of building: *redbrick university* [17, p. 1185]; d) geographical criteria (location): *Cambridge* [17, p.250]. II) Names of the participants of the process of education can be further subdivided on the basis of the people's roles: a) academic staff: *lecturer* [17, p. 804], *reader* [17, p. 1175]; *професор* [14, p. 222]; b) administrative staff and educational authorities: *dean* [17, p. 367]; c) students and pupils: *pupil* [17, p. 1145]; *школяр* [14, p. 705]. Bases for the lexemes naming students and pupils include: 1) period of studies: *fresher* [17, p. 565]; *першокурсник* [14, p. 517]; 2) student's age: *mature student* [17, p. 884]; 3) type of the educational establishment: *preppy* [17, p. 1111]; 4) attitude to studies: *A student* [17, p. 1]; *четвірочник* [14, p. 1408]; 5) period of studies: *dropout* [17, p. 426]; *випускник* [14, p. 204]. III) Components of the educational process: a) types of courses: *major* [17, p. 864]; *спецкурс* [14, p. 1323]; b] forms of assessment: *cloze test* [16, p.310]; *екзамен* [14, p. 224]; c) system of grades: *GPA* [17, p. 617]; *оцінка* [14, p. 388]; d) educational curriculum: *curriculum* [17, p. 3392]; *навчальний план* [14, p. 451];

e) temporal characteristics of education: *academic year* [17, p. 7]; *семестр* [14, p. 527]; f) types of assignments: *homework* [17, p. 685]; *твір* [14, p. 622]. IV) Results of educational process: a) certificates and diplomas: *Dip.H.E.* [17, p. 388]; *атестат* [14, p. 17], *диплом* [14, p. 54]; b) degrees: *MA* [17, p. 987]; *доктор* [14, p. 312]. V) Education authorities: *PTA* [17, p.1104]; VI) Forms of financial support in education: *scholarship* [17, p.1245]; *стипендія* [14, p. 1300]. Following the typology of grades of equivalence, lexical units that verbalize the conceptual field of education in the English and Ukrainian languages belong to one of the following categories: a) full equivalence, for example *dean* [17, p.367] and *декан* [14, p. 22]; b) overlapping, for example *professor* [17, p. 1126], and *професор* [14, p. 677], since in the Ukrainian language the lexeme denotes the academic position of a holder of Doctor of Science degree; c) lack of equivalence, for example *don* [17, p. 400].

Lexical units of various grades of equivalence from each of the thematic groups were run as seed words in SketchEngine and two subcorpora were created: EducationEnglish (507467 tokens) and EducationUkrainian (490320 tokens). Then, the subcorpora (EducationEnglish and EducationUkrainian) were compared against the reference corpora of English language (EngTenTen 19685739271 tokens) and Ukrainian language (UkrTenTen 2194447594 tokens). The reference corpora were not collected for the purpose of the present research but are available to linguists via SketchEngine [18]. Reference corpora are general language collections of written texts obtained from the web and are not topically specific. They were created using similar data collection techniques of web crawling that were applied to obtain EducationEnglish and EducationUkrainian collections. The comparison between corpora relies on keywords in each of the collections. Keywords in EducationEnglish and EducationUkrainian are compared against corresponding reference corpora and sorted according to their log likelihood ratio. The keyword extraction and log likelihood calculation is performed using AntConc 3.4.4w [15]. The lists of keywords are sorted by keyness. The list of top 15 key words in the English and Ukrainian languages is given in Table 1 (conjunctions and prepositions have been removed from the table). Grey cells indicate that the keywords are translation equivalents. Results of the keyword comparison suggest that both subcorpora reflect the concept of education. Therefore the corpus can be regarded as comparable and be used in comparative studies of views English and Ukrainian speakers have about the concept of education.

Table 1.

Key words in English and Ukrainian language corpora

Keyness (English)	Keyword (English)	Keyword (Ukrainian)	Keyness (Ukrainian)
1.997	school	України	0.342
1.578	students	освіти	0.253
1.441	education	навчання	0.208
1.197	learning	університету	0.182
0.785	university	захисту	0.149
0.663	student	роботи	0.143
0.523	schools	дітей	0.130
0.518	year	розвитку	0.129
0.474	course	діяльності	0.128
0.441	college	цивільного	0.128
0.429	research	національного	0.117
0.396	educational	наук	0.114
0.380	first	час	0.112
0.353	work	учнів	0.107
0.350	high	школи	0.104

The conclusions drawn from the study suggest that researchers interested in contrastive analysis of languages can rely on different types of multilingual corpora. Depending on the languages under analysis and research objectives those could include comparable and parallel corpora. Some multilingual corpora can be commercially available. However, not all languages are represented in these types of collections, so linguists might choose to compile their own corpus. Creation of translation corpora might include the human translator intervention and is unlikely to contain as many documents as a comparable corpus. If the goals of the study do not include bilingual lexicon extraction or translation equivalence analysis, comparable corpora might prove more useful for looking into natural non-elicited language data. That being the case, web resources appear to be the first choice of data for corpus creation. Building corpus from web data can be both relatively fast, since it does not include manual text procession or applying for permission to use documents, and allows to access different types of texts to maintain corpus representativeness. One of the drawbacks of using web sources to build a corpus might be the relatively little control linguists have over the types of web pages retrieved for the corpus and the necessity to check for duplicates. What is more, there could be certain restriction on the number of pages accessed daily for the purpose of building a corpus. Building a topic-specific comparable corpus can also be done using the articles from Wikipedia. However, in this case linguists might obtain not a comparable, but a noisy parallel corpus since many articles in encyclopedias are translations. If researchers are interested in building a topic-specific comparable corpus, two techniques for data collection can be used: cross-language information retrieval or clustering. The first approach would be beneficial in collecting data from Wikipedia or multilingual websites such as news agencies reports. The second technique could rely on a set of frequent words or topic-related keywords that are used as query terms if collecting documents. Once the corpus is built, its components should be tested to be comparable. Comparability measures rely on quantitative analysis and statistics that takes into account the number of documents in the corpus. The latter is not always possible to establish, especially in the case of corpora retrieved from the web. Alternative quantitative comparability measures focus on translation equivalence or word frequency similarity. The second technique appears to be more useful for topic-specific comparable corpus analysis since it also allows measuring the topic relevance of key language units. The present study used the procedures outlined above to collect a quasi-comparable (non-aligned) corpus focusing on the topic of education with English and Ukrainian languages in contrast. Using frequency comparability measure it was established that both components of the corpus (in the English and Ukrainian languages) contain keywords related to the topic of education. Furthermore, the obtained keyword lists in two languages contains translation equivalents. Therefore, the corpus on the topic of education with the English and Ukrainian languages in contrast can be used in contrastive studies to analyze the attitudes and views that speakers of the languages under study have about the phenomenon of education. Further research should take into account usage patterns and contexts that key lexical units are used which would provide more insights into the attitudes to the phenomenon of educations verbalized in English and Ukrainian languages.

References:

1. Baroni M. BootCaT: Bootstrapping corpora and terms from the Web / M. Baroni, S. Bernardini // LREC 2004: Proceedings. – Paris. ELRA, 2004. – P. 1313–1316.

2. Costa H. Assessing Comparable Corpora through Distributional Similarity Measures / H. Costa // EXPERT Scientific Technological Workshop. – Malaga, Spain, 2015. – P. 23-32.

3. Fung P. Multi-level bootstrapping for extracting parallel sentences from a quasi-comparable corpus / P. Fung, P. Cheung // 20th international conference on Computational Linguistics: Proceedings, 2004. – P. 1050-1056.

4. Gamallo P. Wikipedia as a multilingual source of comparable corpora / P. Gamallo, I. Gonz´alez// LREC 2010 Workshop on Building and Using Comparable Corpora. – Valeta, Malta, 2010. – P. 19-26.

5. Gelbukh A. Automatic Term Extraction Using Log-Likelihood Based Comparison with General Reference Corpus/ A. Gelbukh, G. Sidorov, E. Lavin-Villa, L. Chanona-Hernandez // NLDB'10 Natural language processing and information systems: Proceedings. – Berlin, Heidelberg: Springer-Verlag, 2010. – P. 248-255.

6. Hajjem M. Building comparable corpora from social networks/ M. Hajjem, M. Trabelsi, Ch. Latiri // 7th Workshop on Building and Using Comparable Corpora 2007. – [Electronic Resource]. − Mode of Access https://comparable.limsi.fr/bucc2014/8.pdf

7. Ke G. Variations on quantitative comparability measures and their evaluations on synthetic French-English comparable corpora / G. Ke, P. Marteau, G. Ménier // 9th Language Resources and Evaluation Conference: Proceedings, 2014. – P. 133-139.

8. Liu S. Termhood-based comparability metrics of comparable corpus in special domain / S. Liu, C. Zhang // Proceedings of the 13th Chinese Conference on Chinese Lexical Semantics. – Berlin, Heidelberg: Springer-Verlag, 2013. – P. 134-144.

9. McEnery T. Parallel and comparable Corpora. The state of the Play / T. McEnery, Z. Xiao // Corpus-based Perspectives in Linguistics. – Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007. – P. 131-145.

10. Saad M. Extracting Comparable Articles from Wikipedia and Measuring their Comparabilities / M. Saad, D. Langlois, K. Smaïli // Social and Behavioral Sciences, 2013. – Vol. 95. – P. 40 - 47.

11. Su Fa. Measuring comparability of documents in non-parallel corpora for efficient extraction of (semi-)parallel translation equivalents/ Fa. Su, B. Babych // Exploiting Synergies between Information Retrieval and Machine Translation and Hybrid Approaches to Machine Translation: Proceedings . – Avignon, France: Association for Computational Linguistics, 2012. – P. 10-19.

12. Talvensaari T. Creating and exploiting a comparable corpus in cross-language information retrieval / T. Talvensaari, J. Laurikkala, K. Järvelin // ACM Transactions on Information Systems, 2007. – [Electronic Resource].– Mode of Access https://tampub.uta.fi/bitstream/handle/10024/66015/creating_and_exploiting_a_comparable_corpus_2007.pdf?sequence=1

Corpora References

13. Великий тлумачний словник української мови: Близько 40000 сл. / [упоряд. Т. В. Ковальова]. Харків: Фоліо, 2005. 767 с.

14. AntConc 3.4.4w. URL: http://www.laurenceanthony.net/software/antconc/

15. EngTenTen. URL: https://www.sketchengine.co.uk/

16. Longman dictionary of contemporary English / [ed. A. Gadsby]. Edinburgh : Pearson Education Limited, 2000. 1675 p.

17. SketchEngine – [Electronic Resource]. – Mode of Access

18. UkrTenTen. URL: https://www.sketchengine.co.uk/