Sadykova A.K.,
Aushakhman A.T.

Electronic text corpus as a tool of a translator

The article analyzes the possibilities to optimize the translation process using modern methods of using the system of language corpus. Corpus Internet technologies can be used effectively in teaching translation, as well as in educational and methodological work of the teacher of higher education institution. The ability to access the meaning of the word in its contextual use is provided by systems of the language corpora, which are seen as a necessary complement to the translator toolkit to enhance and develop the translation competence. The advantages of concordances, translation memory programs, which are to improve the productivity of a translator's work are considered.

Key words: language corpora, concordance, context, translation, translation memory, teaching translation, TM-program.

СадыковаА.К.,
Аушахман А.Т.

Аудармашы құралы ретіндегі мәтіндердің электрондық корпусы

Мақалада тілдер корпусы жүйесін қолдану арқылы аударма үрдісін оптимизациялау мүмкіндіктеріне жан-жақты талдау берілген. Корпустық Ғаламтор-технологиялар аударманы нәтижелі түрде оқытуда, сонымен қатар, жоғары оқу орны оқытушысының оқу-әдістемелік жұмысында қолданылуы мүмкін. Сөздің контекстуалдық қолданылуы аясында сөз мағынасын анықтау әдістерін аудармашылық компетенцияны дамыту мақсатымен аудармашының жұмыс құралдарына қосымша ретінде қажетті болып қарастырылатын тілдер корпусының жүйелері қамтамасыз етеді. Осыған қоса, мақалада аудармашы жұмысының өнімділігін арттырушы конкорданс-бағдарламалар, аударма жадылары бағдарламаларының артықшылықтары да көрсетілген.

Түйін сөздер: тілдер корпусы, конкорданс, контекст, аударма, аударма жадысы, аударма жинақтаушысы, аударманы оқыту, ТМ-бағдарлама.

Садыкова А.К.,
Аушахман А.Т.

Электронный корпус текстов как инструмент переводчика

В статье проводится анализ возможностей оптимизации переводческого процесса за счет современных способов использования системы корпуса языков. Корпусные Интернет-технологии могут быть эффективно использованы в обучении переводу, а также в учебно-методической работе преподавателя высшей школы. Возможность доступа к значению слова в рамках его контекстуального использования предоставляется системами корпуса языков, которые рассматриваются как необходимое дополнение к инструментарию переводчика с целью повышения и развития переводческой компетенции. Указаны преимущества конкордансов, программ памяти переводов, которые заключаются в повышении производительности труда переводчика.

Ключевые слова: корпус языков, конкорданс, контекст, перевод, память переводов, накопитель переводов, обучение переводу, ТМ-программа.

UDC 81’322; 004.934; 004.912

Sadykova A.K.[1], Aushakhman A.T.[2],

[1]cand. of ped. sc., ass. professor,
[2]MA, tutor of KazNU named after al-Farabi, Almaty, Kazakhstan,
e-mail: aida_66@bk.ru; assel.aushakhman@gmail.com

ELECTRONIC TEXT CORPUS AS A TOOL OF A TRANSLATOR

Introduction

In modern society the important role is played by computer technology, which penetrates into the sphere of human activity, forming a global information space. The applicability of computer technology is extremely relevant today. The computerization of the translation process was one of the most important problems from the beginning of information technologies application in science. The dream of creating an automatic machine translation did not leave scientists from the beginning. The introduction of computer tools in the process, is initially focused only on the man, his ability to selecting appropriate option at the level of experience and sense of style, requires special attention to detail and technology. These successes are dependent primarily on the achievements in the study of human thought and verbal communication skills of engineering-linguistic modeling of these processes. Nowadays it is impossible to imagine the work of a translator without a personal computer that is used for the actual translation and to address related problems, for example, to search for background information and learning the terminology.

As it is known, translation is a complex form of human intellectual activity, the transition from the source language to the target language. Translation is a complex multi-faceted phenomenon, some aspects of which may be the subject of study of various sciences. It is a process of intercultural communication in which, basing on an analysis of the translation of the source text a secondary translated text is created to replace the original in the new linguistic and cultural environment.

In the process of translation the translator uses linguistic and extra-linguistic knowledge, and in addition, this process involves two fundamentally different stages: understanding of the text in the original language and the synthesis of the text in the target language. Because of this complexity of the translation process the science about it (Translation Studies) is interdisciplinary in nature and is related to linguistics, literary criticism, cognitive science and cultural anthropology.

The issue of translation has always attracted the attention of linguists and still is one of the most important. In order to overcome some of the difficulties of translation are used advances of comput-

er technology. With the development of computer technologies the number of tools used by the translators have increased many times.

The capabilities of corpus in teaching translation

Resources such as electronic dictionaries, encyclopedias, reference books have replaced the paper-based counterparts. The main advantage of digitized resources is to simplify and accelerate the work with them, which contributes to the rapid development of resources for linguists and translators in particular.

As it is known, the resolution of lexical ambiguity is one of the central tasks of word processing. The objective is to reveal the meaning of words or compound terms in accordance with the context in which they are used. Lexical ambiguity resolution is used to improve the accuracy of the methods of texts classification and clustering, increase translation quality of information retrieval and other applications. In order to solve the problem it is necessary to determine the possible meanings of words and the relationship between these meanings and the context in which the words are used. The linguists compile thesauri, semantic networks and other specialized structures to establish the connection between the meanings. However, the creation of such resources requires a huge effort.

Nowadays, many scientific experiments are carried out in line with the corpus linguistics, whose goal is the study of the process of translation. Prominent in this area are the works of M. Wilkinson, V.N. Shevchuk, N.V. Vladimova and R.K. Koshkin, where the corpus of electronic texts is seen as a means of identifying and addressing the factors that led to the set of «negative effect associated with the «inauthenticity» of the translated discourse.

«Corpora made up of specialised texts can be a useful source of terminology and content information. In the classroom, comparable corpora can be used to confirm translation hypotheses and to suggest possible solutions to actual translation problems related to a specific text. They can also provide a means to investigate similar domains or subdomains across languages. A specialised comparable corpus can offer information about terminology and concepts, and about the attestedness of expressions within a certain context [1].

The corpus of texts is characterized by four basic parameters: 1) it must be of a sufficiently large volume; 2) it should be structured or marked up; 3) the texts that make up a definite corpus should be in electronic form; 4) the concept of «electronic corpus» is, as a rule, special software to work with this corpus. The value of the corpus is defined as follows:

• corpus shows language data in their real environment that allows to explore the lexical and grammatical structure of the language, as well as continuous processes of language changes over a certain period of time;

• corpus is characterized by representative, or a balanced composition of the texts, it can be used to test the search machine, morphology, translation systems, and use it in various linguistic studies;

• corpus is essential to teaching of translation, as it can quickly and efficiently check the features of the use of an unfamiliar word or grammatical form.

Translation is a collection of micro studies that should be conducted by the translators in each case when they are faced with the necessity of solving a problem of translation. The more translators see weaknesses in their translation and the more often they turn to a variety of sources, including corpus of texts, in the search of their information about the use of lexical and grammatical units, the better is the quality of the translation. Additionally, the corpus should be used as a source of further information regarding the domain, in which the translation is implemented. The further development of specialized electronic enclosures texts on various subjects, and their introduction into practice of translation allows rationalizing the work of an interpreter.

Abstract or markup is the main characteristic of the corpus, which distinguishes it from the digital collections, libraries, encyclopedias, widely presented in the modern Internet. Marking up the text is the attribution of certain information for easier analysis of the text [2].

There are different types of markup: 1) the meta-text markup (author, title, date, volume, subject of the text, etc.), which characterizes the text as a whole; 2) structural markup is information about the structure of the text, which allows to distinguish one word from another, highlight border phrases, sentences, text; 3) linguistic markup is attributing certain linguistic units of text information (negative or interrogative sentences, management or contiguity, etc.). The richer and more varied the layout is, the higher is the scientific and educational value of the corpus.

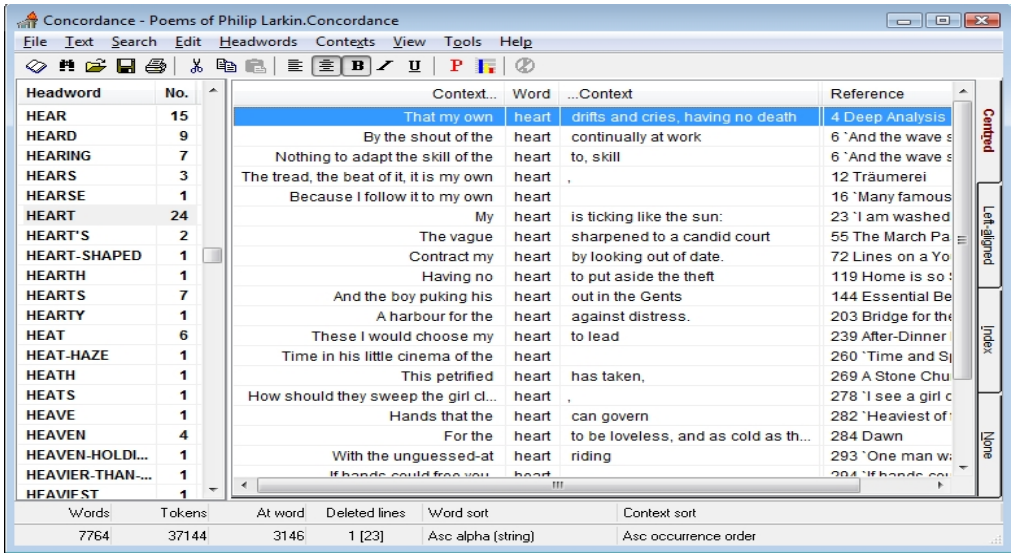
The effectiveness of information retrieval in the corpus depends on special software – the so-called corpus-managers, or concordancer-programs. The space of electronic text corpora has enabled the effective use of electronic concordances, which offer the perspectives of modeling language picture of the

world. Concordance is a specialized language application program, through which the sample is automatically given language units of electronic texts [3, p. 77]. Thus, concordance to a corpus is a list of word tokens, elements of the corpus, with reference to all contexts. The differences between the dictionaries and concordances are in representativeness, orientation on invariant, semantic or grammatical analysis.

The function of concordance can be compared with the function of search in the text editor, but the capabilities of the concordance are much greater: it analyzes not one, but several texts or corpus of electronic texts at one time. Depending on the technical capabilities concordance can provide information about the frequency of use and compatibility of a given language unit, but also enables to access a specific text, in which the example was found.

The comprehensive features of concordancer-programs are limited by linguistic complexity of the search information in the corpus of texts, which depends on how the interpreter is able to put forward the suggestions on possible options for the translation, provide the basis for compiling queries in the search for information.

In other words, concordance is a «program that allows analyzing large amounts of texts in order to detect patterns of use of words and expressions in the language. Concordance-program searches the requested word in the corpus and produces a new window with a few fragments of sentences from different texts where the word or phrase is used. From the results of the corpus search the meaning of the word out of context, and an analysis of its use in the language can be obtained. The search results can be used to clarify the usage and elimination of rules for the use of certain words and phrases in the language, as well as for the study of the grammatical structure of the language» [4, 12-13].



Picture 1 – shows the window of search results from program Concordance 3.3 for the word heart

Some researchers tend to consider concordance not as a program, but as a result of its work. In the interpretation of V.N. Shevchuk« concordance is a vertical list of occurrences of words in alphabetical order in the electronic corpus. Word is supplied together with its left and right environment» [5, 45].

The translator, who has access to the corpus, can see all the examples of words and phrases from the millions of words of text in a few seconds. Not limited (constantly developing) monitor corpora play a huge role in the structure of the dictionary, as they allow to follow the new words, piercing the language, or a pre-existing words to change their meaning or balance their use in accordance with the style. The method of presentation and storage of corpus texts are based on modern computer technology of data storage and processing.

The corpus of texts can be considered to be one of those means, the use of which in certain cases should be referred to a necessary condition for analysis of linguistic phenomena. For example, modern electronic corpus contains hundreds of millions of word tokens, which allows us to speak about their viability in terms of the language competence.

N.V. Vladimov notes that the «basic procedures that are available to the researcher in the analysis of corpus include:

• search for the specified words, phrases in the corpus;

• display of search results, taking into account the features of a specific field;

• counting the number of examples of use of the word in the corpus;

• sorting of search results based on the required parameters.

All these procedures are carried out quickly and accurately by a computer program of concordance compilation (searching for equivalents), which allows researchers to quickly and accurately find what they need» [6, 26].

At this point in the wide popularity acquires the possibility of using the Internet as the corpora of the texts. This became possible after the on set of online web-concordancers, the so-called «Web as well Corpus». This resource has a number of disadvantages, and it is less effective than thematic corpora working in anautonomous regime. However, given the shortage of time and lack of specialized corpora ready online-concordancers can be a source for reliable linguistic information.

Virtual (specialized) corpus is a vast in terms of volume on specific topics, specially composed to find certain linguistic information text selection for the translators. The texts are taken from various sources (periodicals, encyclopedias, the Internet) in a strictly defined category and always presented in an electronic form. We can say that virtual corpora produced by a translator on specific topics may help him in the following cases:

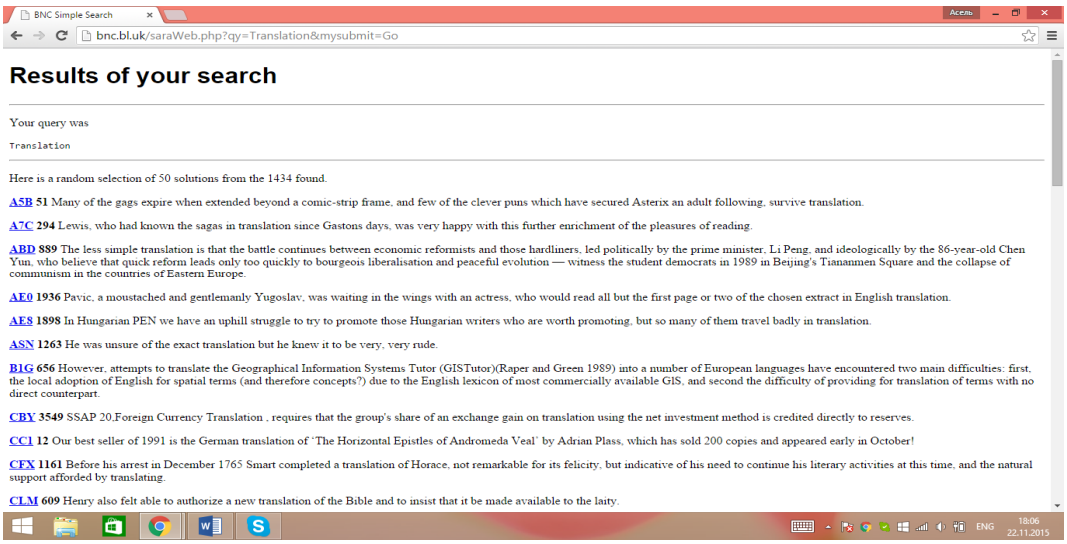
• to define the lexical and grammatical compatibility of words;

• to select from several options lexical equivalent of the original word, offered in different dictionaries or the Internet;

• to validate the decisions intuitively selected by the translator;

• to find additional encyclopedic information on the subject;

• to find terminological doublets, antonyms, holonyms, meronyms, identifying names and definitions of terms [5, 52-57].



Picture 2 – shows an example of the search results of the word translation in the Web-based version of «British National Corpus.»

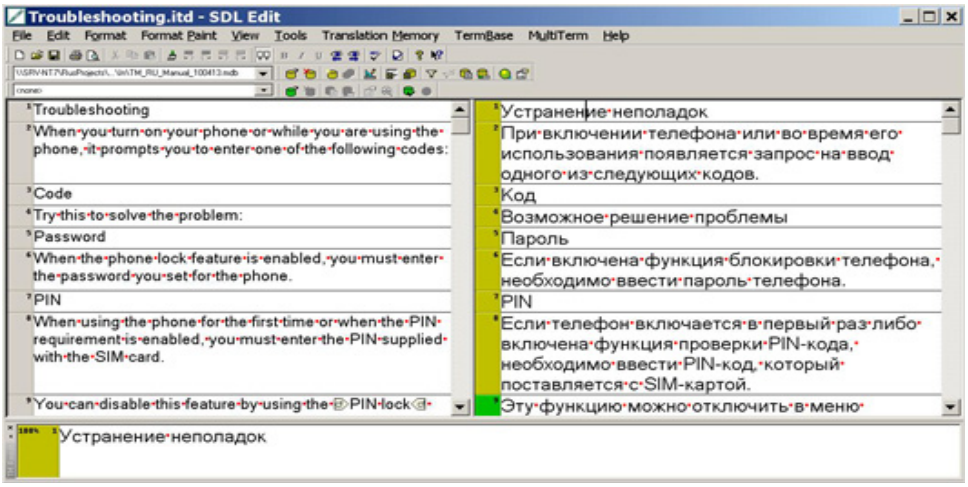
Many researchers in the field of corpus linguistics point out that the corpus gives the translator the opportunity to actually navigate the language and solve a certain number of linguistic and extra-linguistic problems in the translation process. V.N. Shevchuk notes, «This is a powerful and reliable electronic resource, in practice replaces the actual native speaker, and becomes the so-called «virtual native speaker», the use of which, no doubt, contributes to the quality of translation. The corpus gives a clear idea of the lexical, grammatical, stylistic, spelling and punctuation rules, operating in a modern language» [5].

Supporting programs for translation or CAT (computer assisted translation) are computer programs that automate the routine operations of

the translator, freeing him time for intellectual tasks. Unlike machine translation, where human intervention is required only to download the dictionary and to edit the resulting output of the text, the translation process using CAT-programs is constantly and entirely under the control of human. The automation means avoid the multiple translation of text fragments that have been translated before and store the translation variants in the so-called Translation Memory or TM. For example, if translator has already translated the phrase «actuator casing cap screws and hex nuts», then the next time he will meet this phrase in the text, the program will offer a translation automatically. CAT-programs allow the translator to accumulate in TM bilingual word pairs «original – translation.» During the work on the texts, similar in genre and subject matter, the translation process is accelerated due to the replenishment of TM [7, 57-62].

The segment of the original and compared with it segment of translation represent a translation unit (TU). The role of the segment is usually implemented by the sentence. TM-programs are distinguished by several criteria. Thus, by the technical realization are allocated local TM-programs and the programs available in the online regime. When using the online versions (for example, Google Translator Toolkit and Word fast Anywhere) the work is conducted in the browser window.

According to the second criterion, the additional functional possibilities: built-in programs to provide access to terminology databases, complement the translation memory (TM). This ensures consistency of terminology in the translation, to ensure compliance with the language policy in the framework of a single enterprise or domain. In order to build a translation memory the segments of original text and translation must be compared for their further entering into a database program. Thus, TM-programs are provided with a function of text segments grading (usually sentences). Furthermore, directly in the translation process for the resolution of lexical units disambiguate meaning may be required the viewing of a word in the context. TM-programs, as a rule, provide an opportunity to build a concordance of a word: to find a list of all occurrences of a word in the context, at the same time presenting ways of translating the lexical unit in each case.



Picture 3 – presents the active window of the TM-program

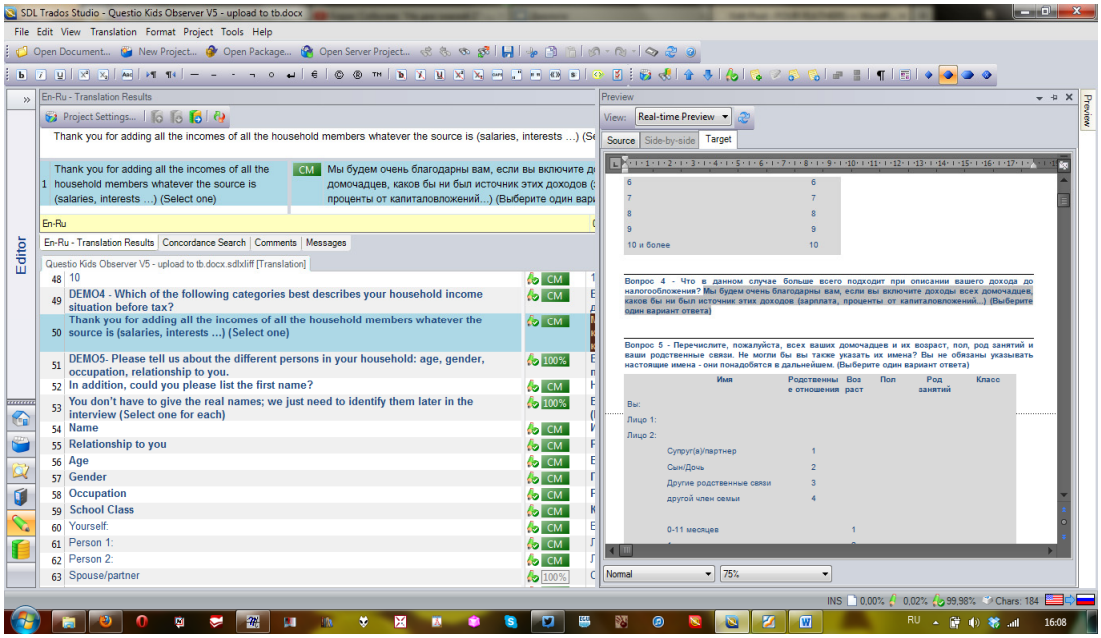
According to O.I. Babina, P.G. Osminin, the actual process of translation with the use of translation memory involves several steps.

1. Preparation: At this stage should be solved the problem of creating a new translation memory, or make a selection of the translation memory to be used for the translation of the text, to make the necessary adjustments (for example, to focus on ways of translating the term corresponding to the previously created projects) and adjustments (e.g., replacing a specific term in all translation units in accordance with a certain project) of the selected translation memory.

2. Translation: provisional translation (pre-translation) is an automatic replacement of the exact segments of the source text to their equivalents in the target language. The segments of the text

that were not found in the database are translated manually or, if available, using the machine translation system. The TM-commercial programs, as a rule, have in-built systems of terminology extraction (e.g., Extract, SDL MultiTerm, PROMT TerM, MonoConc Pro, Simple Concordance Program, and others.). «Raw» text, including automatically replaced terminology and text segments in the final stage is subjected to post-editing.

3. Quality control: includes a formal check for completeness of the translation, grammatical accuracy, and correct translation of the relevant terminology, which can be carried out by a translator, (possibly) by the customer [8].



Picture 4 – shows the work with the TM-memory program TRADOS.

The types of tasks performed by students in the classroom for the translation can be diverse and reflect the specifics of the real work of a translator. These tasks may include:

• task on compiling a glossary of the text (the students receive a text of particular subject area and make up for it a translation glossary);

• task on the alignment of the text and its translation;

• task on a preliminary analysis of the efficiency and concordance of TM;

• task on translation using a concordance and TM;

• task on comparing translations, implemented with the use of the corpus and TM.

Conclusions and methodological recommendations

Concluding the above should be noted that almost all the shortcomings of concordances and tasks based on them are associated with considerable expenditure of time. Thus, students and teachers need to acquire special skills that take time and effort. In addition, currently there is practically no finished material of this type. This means that the teacher should develop special tasks and render a huge volume of the material.

However, the role of the corpus in teaching translation is great. In fact, all of the major mistakes of students take place because of the fact that they did not have a good example of using particular words, expressions, and are only theoretically aware of the structure of different text styles. Give the students the opportunity to understand the application of a rule in practice is realistic and necessary. It should also be noted that the knowledge obtained in the course of their own research, is mostly deposited in the memory than ready conclusions. Students will gladly share the results of their research, and promotion of this work will be the best stimulus for learning a foreign language.

The tasks prepared in advance to a definite corpus can be used at the lessons of foreign language and

practice of translation: ensure the use and translation of prepositions (for example, *about*); analyze the complex sentences, interrogative sentences; analyze the differences in the use of words *tell/say, listen/hear,* different terms; semantization of words (for example, *make* – analyze the use of verbs in different contexts); find international words and tell if their meaning is the same in the native and the target language, and etc.

Conclusion

In the objectives of teaching translation corpus of texts can be seen as abstract information and provide samples of professional translation in the study of methods and techniques of translation. An analysis of text corpora, the methods of corpus linguistics and achievements are a promising direction in the field of teaching foreign languages and translation. The world practice of development in this field proves their effectiveness. The toolkit of working with words and expressions the proposed by the corpus creates additional opportunities for saving the translator's time in finding the equivalents and contributes to the accumulation of knowledge of communicative-heuristic character, allowing the translator to cognize the language in a complex way, with the context, that is an effective strategy to address translation problems.

We believe that the corpus is more convenient and reliable means in comparison with dictionaries for several reasons. First, the corpus of texts is not a set structures as traditional dictionaries, but is constantly updated database. With it translator is able to keep up with the latest trends in the development of language-based analysis of the use of a word in the corpus. Secondly, the corpus is a vast source of wordtokens than a dictionary due to the fact that it is much larger in volume, and the information about the word, which can be obtained from the corpus, is more objective and accurate. Thirdly, work with the corpusis much easier than with the dictionary. The corpus is placed on a machine (computer), which is currently part of work place of a translator. By means of simple manipulations of the keyboard and «mouse» translator in a few seconds is able to get the needed linguistic information.

Литература

1 Pearson, J. Teaching Terminology using Electronic Resources, S. Botley, J. Glass, T. McEnery and A. Wilson (Eds), Proceedings of Teaching and Language Corpora 1996. UCREL technical Papers, Lancaster, UCREL. – P. 203-216.

2 Чепик Е.Ю. Политическое слово в структуре электронного словаря // Культура народов причерноморья. – № 69. – 2005. – С. 205-210.

3 Бовтенко М.А. Компьютерная лингводидактика: Учебное пособие / А. Бовтенко. – М.: Флинта: Наука, 2005. – 216 с.

4 Сысоев П.В. Иностранные языки в школе. – М.: ООО «Методическая мозаика», 2010. – Вып. 4. – С. 12-13.

5 Шевчук В.Н. Электронные ресурсы переводчика: Справочные материалы для начинающего переводчика. – М.: Либрайт, 2010. – С. 45-57.

6 Владимов Н.В. Корпусный подход к решению переводческих проблем: На материале письменных переводов с русского языка на английский: диссертация ... кандидата филологических наук: 10.02.19. – Москва, 2005. – С. 26.

7 Грабовский В.Н. Технология TranslationMemory // Мосты. Журнал переводчиков. – 2004. – № 2. – С. 57-62.

8 Бабина О.И., Осминин П.Г. Память переводов в обучении переводчиков. – 2013. – Т. 5. – № 3.

References

1 Pearson, J. Teaching Terminology using Electronic Resources, S. Botley, J. Glass, T. McEnery and A. Wilson (Eds), Proceedings of Teaching and Language Corpora 1996. UCREL technical Papers, Lancaster, UCREL. – P. 203-216.

2 Chepik E.YU. Politicheskoe slovo v strukture ehlektronnogo slovarya // Kul'tura narodov prichernomor'ya. – № 69. – 2005. – S. 205-210.

3 Bovtenko M.A. Komp'yuternaya lingvodidaktika: Uchebnoe posobie / A. Bovtenko. – M.: Flinta: Nauka, 2005. – 216 s.

4 Sysoev P.V. Inostrannye yazyki v shkole. – M.: OOO «Metodicheskaya mozaika», 2010. – Vyp. 4. – S. 12-13.

5 Shevchuk V.N. EHlektronnye resursy perevodchika: Spravochnye materialy dlya nachinayushchego perevodchika. – M.: Librajt, 2010. – S. 45-57.

6 Vladimov N.V. Korpusnyj podhod k resheniyu perevodcheskih problem: Na materiale pis'mennyh perevodov s russkogo yazyka na anglijskij: dissertaciya ... kandidata filologicheskih nauk: 10.02.19. – Moskva, 2005. – S. 26.

7 Grabovskij V.N. Tekhnologiya TranslationMemory // Mosty. ZHurnal perevodchikov. – 2004. – № 2. – C. 57-62.

8 Babina O.I., Osminin P.G. Pamyat' perevodov v obuchenii perevodchikov. – 2013. – T. 5. – № 3.