



UDK:811.523

Liliya YUSUFOVA

Jizzakh State Pedagogical University

Department of practical English language

E-mail: yusufova.liliya.zaamin@mail.ru

Under the review of Gulyamova Mavluda Hatamovna PhD in pedagogy of UzSWLU

FORMAL-FUNCTIONAL FEATURES OF VERB PHRASES BASED ON THE ANALYSIS OF PARALLEL CORPUS (ON THE EXAMPLE OF ENGLISH-UZBEK TEXTS)

Annotation

In the article there are considered about the parallel corpus, its structure, composition and possibilities in the languages. The first studies on parallel corpus creation in linguistics were analyzed. In particular, it is thought that the material contained in parallel corpora is divided into bilingual and multilingual parallel corpora according to the language, bilingual corpora consist of translation and parallel corpus, and multilingual corpora are made up of several translations of the original text.

Key words: Linguistic corpus, national corpus, parallel corpus, parallel text (bixtext), fragment, syntactic unit, monolingual corpus, bilingual corpus, multilingual corpus.

ПАРАЛЛЕЛ КОРПУС ТАХЛИЛИ АСОСИДА ФЕЪЛИИ ФРАЗЕМАЛАРНИНГ ФОРМАЛ-ФУНКЦИОНАЛ ХУСУСИЯТЛАРИ (ИНГЛИЗ-ЎЗБЕК МАТНЛАР МИСОЛИДА)

Аннотация

Мақолада параллел корпус, унинг тузилиши, таркиби ва тиллардаги имкониятлари ҳақида фикр юритилади. Тилшуносликда параллел корпус яратилиш бўйича дастлабки тадқиқотлар таҳлил қилинди. Хусусан, параллел корпус таркибидagi материал тилга кўра икки тилли ва кўп тилли параллел корпусларга бўлинади, икки тилли корпус таржима ва параллел корпусдан, кўп тилли корпус эса асл матннинг бир неча таржималаридан иборат деб ҳисобланади.

Калит сўзлар: Лингвистик корпус, миллий корпус, параллел корпус, параллел матн (битекст), фрагмент, синтактик бирлик, бир тилли корпус, икки тилли корпус, кўп тилли корпус.

ФОРМАЛЬНО-ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ ГЛАГОЛЬНЫХ СЛОВСОЧЕТАНИЙ НА ОСНОВЕ АНАЛИЗА ПАРАЛЛЕЛЬНОГО КОРПУСА (НА ПРИМЕРЕ АНГЛО-УЗБЕКСКИХ ТЕКСТОВ)

Аннотация

В статье рассматривается параллельный корпус, его структура, состав и возможности в языках. Проанализированы первые исследования по созданию параллельных корпусов в лингвистике. В частности, считается, что материал, содержащийся в параллельных корпусах, делится на двуязычные и многоязычные параллельные корпуса в зависимости от языка, двуязычные корпуса состоят из перевода и параллельного корпуса, а многоязычные корпуса состоят из нескольких переводов исходного текста.

Ключевые слова: Лингвистический корпус, национальный корпус, параллельный корпус, параллельный текст (битекст), фрагмент, синтаксическая единица, одноязычный корпус, двуязычный корпус, многоязычный корпус.

Introduction. Corpus linguistics is an actively developing field of linguistic research that provides access to a large amount of empirical material for other branches of linguistics. In corpus linguistics, corpora of parallel texts play an important role. At the same time, due to the increasing importance of multilingualism in linguistic fields, the use of corpora of parallel texts is one of the promising technologies with global markets and worldwide information exchange, and its future is very bright. With the help of parallel corpora, it is possible to know the variants of a word, sentence, paragraph, supersyntactic whole in different languages. Parallel corpora are an important reality in today's era of widespread intercultural communication. Through parallel corpora, it will be possible to identify universals in different language environments and cultures, as well as specific mental characteristics of languages, realia and lacunar units. The corpus of parallel texts also serves for the development of automatic translation, ensures the development of computer lexicography. Using the corpus of parallel texts, special concordance programs are developed and it is possible to create dictionaries of various specializations.

Literature review. Parallel corpora are often created from texts used in multilingual communities such as the

United Nations, the European Union, and officially bilingual countries such as Canada. The main disadvantage of the parallel corpus is that the texts are not free from the influence of the source language, translation errors and the individual style of the translator. Similar problems of disparity are not absent in monolingual corpora. In any case, the solution is the same: the larger and more diverse the corpus, the less the impact of individual text. An early example of a corpus of parallel texts was found in 1799 in the Nile Delta near Rosetta, dating back to 196 BC.

According to D.O. Dobrovolsky, a corpus of parallel texts is useful in not only translation theory, bilingual lexicography, and other linguistic fields related to language comparison, but also for a more detailed description of the semantics of the lexical units of each of the languages being compared[1]. Currently, the parallel text corpus is defined as follows, all of which exhibit the characteristics of a parallel corpus:

A corpus of parallel texts is an electronic representation of works of art, manuals, mass media, various documents in two or more languages.

A parallel corpus is a special type of corpus in which a text in one language compares a translation of this text in

another language, or vice versa, a text in a foreign language compares its translation in that language. Using a process called equalization, alignment is established between the original and translated parts of the text. A balanced parallel corpus is a powerful tool for scholarly research, including translation theory and practice.

The corpus of parallel texts is formed for scientific and practical purposes (in particular, for teaching foreign languages). By its structure, it is a collection of texts in the source language, considered as the translation of one or more texts in the source language into the target languages.

A parallel corpus or a translation corpus is a corpus consisting of texts in one language and their translation into another language or languages. Such parallel pairs or lines are called bitext. To create a parallel corpus, it is not enough to have the original and its translation. It will be necessary to ensure that the passage corresponding to the original is found in the translation. For this purpose, the alignment process is used, as a result of which identical parts of parallel texts are compared with each other.

In Uzbek linguistics, the first researches on creating a parallel corpus can be observed among the works of B. Mengliyev, R. Karimov. They explain parallel corpora as follows: "The corpora are divided into monolingual and multilingual corpora types according to the language of the material contained in the corpora. Bilingual or multilingual corpora are (1) parallel or translation corpora; (2) has the appearance of a comparative corpus".

According to the structure of the experts, the arrangement of the text, the adaptation of the units, there are several types of parallel corpus:

- 1) one-way (text translated from English to Russian);
- 2) two-way (text translated from English to Russian and from Russian to English (back translation));
- 3) multilingual (text translated from English to Russian, German, French). According to the language of the material contained in the corpora, they are initially divided into monolingual and multilingual corpora types.

Bilingual or multilingual corpora have the following appearance:

- 1) parallel or translation corpora;
- 2) comparative corpus. Regardless of the purpose of parallel housing, the units in it should be adapted. Adaptation is carried out on the basis of bitext software, as mentioned above.

Research Methodology. The parallel corpus of texts is the most suitable for translation research, it allows to achieve high representativeness and reliability in conducting research as a modern tool for analysis of translation alternatives.

Corpus specialists (mainly translators) have always been interested in creating multilingual corpora. Since the first generation of corpora, bilingual corpora for English, Finnish, French, German, Greek, Norwegian, Spanish, Swedish, and other languages have begun to emerge. Such a corpus is also called bitexts. Brian Harris originally coined the term bitext in 1988 to refer to documents and their translations into other languages for use in translation studies. Since then, bitext has attracted a lot of interest among the general public, taking into account many other programs. Nowadays, it is customary to use the term bitext to refer not only to original documents and their direct translations, but also to more parallel sources. Such sources may contain different translations from the common source, sometimes even using other intermediate languages used to create the final translations. In computational linguistics, the term parallel text is often used synonymously with the term bitext. Unfortunately, this leads to confusion with translation theory and terminology as Jean Veronis (2010) points out.

In translation studies, the term parallel text usually refers to a collection of texts in the same field, but does not include translations of the same documents. However, in computational linguistics, such corpora are usually referred to as (bilingual) comparable corpora. After several decades of research on bitexts in computational linguistics, the term parallel corpus has now been coined to name a collection of bitexts.

According to the criterion of parallelism, experts in the field divide corpora into monolingual, bilingual and multilingual types. Dialects, language variants are compared in monolingual corpora. Another embodiment of a bilingual, multilingual corpus contains both the original text and the translated text. This type of corpus serves as an important, useful resource in comparative cross-sectional research, translation theory, and computer translation studies. Bilingual and multilingual corpora can be divided into two main types:

1) corpora representing many original texts written in any source language and translations of these original texts into one or more other languages;

2) corpora combining texts on the same topic, independently written in two or more languages; Such corpora help with terminology and are often used by translators. Both types of corpora are used for the development of effective methods of translation, including the creation of bilingual and multilingual terminological dictionaries for machine translation, as well as for the comparative study of languages (in the field of lexicology, grammar, stylistics, translation studies, etc.).

Analysis and results. When translating expressions that deviate from the norms of modern usage, an additional difficulty arises. Dostoevsky's contemporaries apparently perceived the texts of his works differently than we do today. Much of what the modern reader assesses as a deviation from the norm was conventionally acceptable (cf., for example, such phrases as take measures, understand one's own price, very wonderful).

In principle, among non-standard expressions, one should distinguish between linguistic structures that corresponded to the norms of that time, and structures perceived by Dostoevsky's contemporaries as significant deviations from the usage. The latter must necessarily be taken into account in the translation, while the former can be ignored. In theory, such considerations should be taken into account. However, for the consistent implementation of translation strategies focused on such goals, in-depth preliminary research is needed. One should know which elements of the text Dostoevsky's contemporaries perceived as standard and neutral, which evoked additional associations, and which (perhaps being standard today) belonged to the area of non-standard use of the language. Next, the corresponding parallels must be found in the L2 language. Such a task is so difficult to accomplish that it can be left out of the discussion. Comparing the two basic translation strategies, it is important to keep in mind that the differences between them are gradual. It is difficult to imagine a translation of a well-known literary text of the past, focusing exclusively on "communicative" or "philological" settings.

Nevertheless, it is quite possible to trace certain trends, which in turn allows us to put forward meaningful hypotheses about the influence of the cultural and historical context on the preference for one or another translation strategy. To do this, it is first necessary to propose some linguistic parameters relevant to the corresponding types of translation. In other words, before trying to answer the question of interest to both literary scholars and specialists in the theory of translation, why the same text was translated in different periods in such different ways and how this is connected with certain general cultural attitudes, it is

necessary to develop criteria that allow us to attribute each specific translation to the corresponding type. At the same time, it is necessary to have a tool that allows you to quickly search for the information you are looking for and simultaneously access several translations in their comparison with the original, therefore, a prerequisite for solving such problems and obtaining empirically reliable results is the presence of parallel text corpora, including, along with the original text, at least, its two translations to L2.

It seems possible (at least as a working hypothesis) to determine for each literary text some "reference points" - structures that, by their nature, require non-trivial decisions from the translator. In such cases, the translator must choose between the "alienated" version (i.e., in principle understandable to the reader, but violating the usual L2

language norms) and the "distorting" version (observing the L2 conventional norms, but deviating from the original).

Conclusion/Recommendations. We propose the following "reference points" as diagnostic benchmarks in assessing the dominant translation strategy:

1. non-standard use of lexical units, violation of existing compatibility standards;
2. irregular syntactic constructions;
3. lexical units, especially frequently used by this author (favorite words);
4. collocations and typical (mostly non-free) phrases for a given author;
5. idioms and conventional metaphors;
6. author's metaphors;
7. culturally specific elements, including personal names, titles, forms of address, etc.

REFERENCES

1. Dobrovolsky D.O. Corpus of parallel texts in the study of lexical semantics [Text] / D.O. Dobrovolsky // Computational Linguistics and Intelligent Technologies: Proceedings of the Intern. Conference "Dialogue 2004". - M.: Nauka, 2004.
2. Veronis J. (ed.), Parallel Text Processing Kluwer Academic Publishers. 2000. p. 2.