



IJTIMOYIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNALI

ISSN 3060-5059



VOL.3 № 5

2026

TARJIMASHUNOSLIKDA KORPUS LINGVISTIKASINING METODOLOGIK XUSUSIYATLARI

Zahriddin Xaitqulov Xazratqul o'g'li
O'zbekiston davlat jahon tillari universiteti, dotsent (PhD)
Abror Shaymardanov Hamid o'g'li
"Cyber University" davlat universiteti, katta o'qituvchi

Annotatsiya

Hozirgi davrda korpus lingvistikasi turli sohalarda ko'p tilli ma'lumotlarni yig'ish, saqlash va tahlil qilish imkoniyatlarini kengaytirayotgan muhim ilmiy yo'nalishlardan biri hisoblanadi. Shu bois korpus lingvistikasining tarjima jarayonidagi o'rni ham tobora ortib, tarjima sifati va uning barqarorligini ta'minlashda muhim vosita sifatida namoyon bo'lmoqda. Bundan tashqari, zamonaviy tarjima dasturlari manba tildagi leksik birliklar, iboralar, frazeologizmlar hamda katta hajmdagi matnlarni tarjima tiliga o'girishda samarali texnologik yordamchi vosita vazifasini bajarmoqda. Bugungi kunda ko'plab elektron tarjima dasturlari o'zining tarjima xotirasi bazasini shakllantirmoqda va uni ko'p tilli elektron manbalar asosida boyitib bormoqda. Bunday jarayonda korpusga asoslangan tahlil usullari tarjima faoliyatini sezilarli darajada yengillashtiradi. Tarjima tizimlarini yaratishdagi eng muhim va murakkab bosqichlardan biri tarjima xotirasini shakllantirish masalasi hisoblanadi. Aynan korpus lingvistikasi ushbu jarayonni soddalashtirish va samaradorligini oshirishga xizmat qiladi. Bu esa korpus texnologiyalarining tilshunoslik va tarjimashunoslikdagi asosiy afzalliklaridan biri sanaladi. Shu sababli korpus yoki parallel korpus asosida avtomatlashtirilgan tarjima tizimlarini yaratish metodologiyasini o'rganish va ilmiy jihatdan tahlil qilish muhim ahamiyat kasb etadi. Mazkur tadqiqot tarjimashunoslikda korpus lingvistikasining funksional imkoniyatlari hamda metodologik asoslarini yoritishga bag'ishlanadi.

Kalit so'zlar: Avtomatlashtirilgan tarjima, korpus, parallel korpus, metodologiya, korpusga asoslangan tarjimashunoslik (KAT), tadqiqot, tarjima xususiyatlari, korpuslar, kompyuter yordamidagi tarjima (KYT), neyron mashina tarjimasi (NMT), tahlil.

КОРПУСНАЯ ЛИНГВИСТИКА КАК МЕТОДОЛОГИЧЕСКИЙ ИНСТРУМЕНТ В ПЕРЕВОДОВЕДЕНИИ

Захриддин Хаиткулов Хазраткул угли
Узбекский государственный университет мировых языков, доцент (PhD)
Аброр Шаймарданов Хамид угли
Государственный университет «Cyber University», старший преподаватель

Аннотация

В настоящее время корпусная лингвистика является одним из наиболее востребованных научных направлений, обеспечивающих сбор, систематизацию и обработку большого объема многоязычных данных в различных сферах. В связи с этим значение корпусной лингвистики в области перевода постепенно возрастает, способствуя повышению качества и устойчивости переводческой деятельности. Наряду с этим современные переводческие программы становятся эффективным технологическим средством для перевода лексических единиц, словосочетаний, фразеологических выражений, а также крупных текстовых массивов с языка-источника на язык перевода. Сегодня разрабатывается значительное количество электронных переводческих систем, которые формируют и расширяют базы переводческой памяти (translation memory), наполняя их многоязычными электронными ресурсами. В данном процессе корпусно-ориентированный анализ играет важную роль, значительно облегчая переводческую работу. Одной из наиболее сложных задач при создании автоматизированных переводческих программ является формирование качественной переводческой памяти. Именно корпусная лингвистика способствует упрощению и оптимизации данного процесса. Это считается одним из ключевых преимуществ корпусных технологий в лингвистике и переводоведении. Поэтому исследование и осмысление методологии создания автоматизированных систем перевода на основе корпусов и параллельных корпусов приобретает особую научную значимость. Настоящее исследование посвящено изучению функциональных возможностей корпусной лингвистики и её методологических основ в переводоведении.

Ключевые слова. Автоматизированный перевод, корпус, параллельный корпус,

методология, корпусно-ориентированное переводоведение (CBTS), исследование, особенности перевода, корпусы, компьютеризированный перевод (CAT), нейронный машинный перевод (NMT), анализ.

CORPUS LINGUISTICS AS A METHODOLOGICAL TOOL IN TRANSLATION STUDIES

Zahriddin Xaitqulov Xazratqul o'g'li

Uzbekistan State World Languages University, Associate Professor (PhD)

Abror Shaymardanov Hamid o'g'li

Cyber University State University, Senior Lecturer

Abstract

Nowadays corpus linguistics is in the top of all sphere which is collecting and providing a huge number of data in multilingual dimensions. So, the role of corpus linguistics also increasing the quality and sustainability in translation field. On the other hand, translation tools are becoming one of the useful program that assist to translate lexical words, phrases, idioms and even a huge number of texts from source language to target language technology area. A number of e-translation software is creating and developing its translation memory and is being filled in e-library with multilingual sources. In this situation corpus-based analysis plays a vital role to make job easier. The first challenging task for creating translation programs is its translation memory. Corpus linguistics is reducing to create translation memory. This one of the advantage side of corpus in linguistics and translation fields. For this reason, it is crucial to study and comprehend methodology of creating automated translation machine via corpus or parallel corpus. Thus following research deals with the function of corpus linguistics and its methodology in translation studies.

Keywords: Automated translation, corpus, parallel corpus, methodology, corpus-based translation studies (CBTS), research, translation features, corpora, computer-assisted translation (CAT), Neural machine translation (NMT), investigation.

The general strategy and justification for the research process are referred to as methodology. The methodological framework of techniques, instruments, processes, and strategies used to gather, examine, and evaluate data for a study is known as methodology in academic writing. It describes how the study is carried out and the rationale for the selection of particular techniques. Methodology in corpus-based translation research encompasses the kind of corpus chosen, data collection methods, data processing tools, and analytical approaches. Methodology is the larger framework that explains and arranges various tactics into a logical plan, whereas strategy refers to the particular actions or approaches used in the research. **Creswell (2014)** asserts that methodology encompasses the logical flow that directs the entire investigation as well as the philosophical assumptions underlying the research [3; 90-p.].

The incorporation of corpus-based methodologies into translation studies has fundamentally reshaped contemporary approaches to the analysis, interpretation, and pedagogy of translation. Rooted in empirical and data-driven research, this approach employs corpora—systematically organized and electronically accessible collections of authentic texts—to investigate translation phenomena through objective, consistent, and reproducible methods.

In contrast to traditional translation theories, which frequently depend on intuitive judgments, isolated examples, or prescriptive assumptions, Corpus-Based Translation Studies (CBTS) provide a descriptive and analytical framework for examining authentic language use in translation processes. Such an approach enables scholars to identify recurrent translational patterns, language regularities, stylistic tendencies, and context-dependent semantic or syntactic shifts across extensive textual datasets.

Moreover, corpus-oriented methodologies facilitate a deeper understanding of the linguistic and pragmatic mechanisms underlying translation by allowing researchers to examine naturally occurring data within real communicative contexts. Consequently, CBTS moves translation research beyond fragmented observations and subjective interpretations toward more systematic, evidence-based, and generalizable conclusions regarding translation strategies, translator behavior, and intercultural language mediation.

One of the principal reasons for adopting corpus-based methodologies in translation studies is their ability to reveal recurrent translational tendencies commonly referred to as translation universals. These include phenomena such as simplification, explicitation, normalization, and leveling out, which are frequently observed across different languages, text types, and translational contexts irrespective of

individual translator preferences. The identification of such regularities enables researchers to investigate the distinctive linguistic nature of translated discourse in comparison with originally produced, non-translated texts.

Besides, corpus-oriented research provides extensive opportunities for both cross-linguistic and intra-linguistic analysis through the application of parallel and comparable corpora. Parallel corpora, consisting of source texts alongside their translated equivalents, allow scholars to examine correspondences between languages at grammatical, lexical, semantic, and pragmatic levels. Comparable corpora, by contrast, facilitate the comparison of translated and non-translated texts within the same language, thereby offering insights into stylistic and structural features characteristic of translated language. Such analyses contribute to a deeper understanding of how equivalence, variation, adaptation, and meaning transfer are realized throughout the translation process.

In addition to its theoretical significance, the corpus-based approach possesses considerable practical value in both translator education and professional translation practice. By consulting authentic corpus data, translation trainees and practitioners are able to observe how experienced translators resolve lexical ambiguities, interpret culture-specific units, maintain terminological consistency, and reproduce stylistic nuances in diverse communicative contexts. As a result, corpus-informed methodologies promote the development of more accurate, context-sensitive, and evidence-based translation strategies.

Furthermore, corpus resources constitute an essential foundation for the advancement of modern language technologies. They play a crucial role in the development and optimization of translation memory systems, terminology management tools, automated term extraction software, and neural machine translation engines. Consequently, corpus-based technologies contribute significantly to improving translation quality, consistency, productivity, and overall efficiency within contemporary multilingual communication environments.

Literature Review

There are a number of scientists who have worked on research related to automated translation and corpus-based studies. We may mention some of these scholars: M. Baker, S. Laviosa, A. Chesterman, F. Zanettin, G. Wang, P. Corpas, and others.

M. Baker introduced the idea of using corpora to investigate translation universals and patterns. Baker emphasized that corpora allow scholars to analyze language as it is actually used, rather than relying on introspective or anecdotal methods. According to her, corpus linguistics can enable a more objective and systematic exploration of translated texts, uncovering general tendencies in translation practices across languages and genres [1; 233–250-p.].

S. Laviosa notes that the corpus-based paradigm fosters a shift from speculative theorizing to empirical observation. She argues that “translation studies should be viewed as a branch of empirical and descriptive language study”, grounded in the real usage of language in translation contexts [4; 76-p.]. Translated texts tend to use simpler syntactic and lexical structures than original texts. S. Laviosa found that translated English texts displayed reduced lexical variety and sentence complexity compared to non-translated English texts [5; 89-p.].

Translators often make implicit information in the source text more explicit in the target text. B. Blum-Kulka first proposed this hypothesis, which was later supported through corpus-based studies [2; 17–35-p.]. A. Olohan suggests that corpora can be used to support exercises in comparative translation, terminology research, and stylistic analysis [8; 67-p.]. A. Munday, Bernardini, and Wallmach have brought the debate to the fore by suggesting how corpus data can provide insights into the grammatical and cohesive structures of translated texts [7; 34-p.].

Methods

Translation studies have employed multiple types of methods to collect data in different languages in recent years. Some of the most important methods used for translating texts from a source language into a target language are direct translation, word-for-word translation, contextual translation, statistical translation, rule-based translation, and neural translation.

The above-mentioned translation methods have a long history extending from the past to the present period. The founders and contributors to these automated translation methods include J. Hutchins, M. Baker, S. Tiogo, and others.

Results

One of the most extensively studied areas within CBTS concerns translation universals. M. Baker proposed that translated texts exhibit certain recurring linguistic tendencies regardless of language pair or

translator identity. These include:

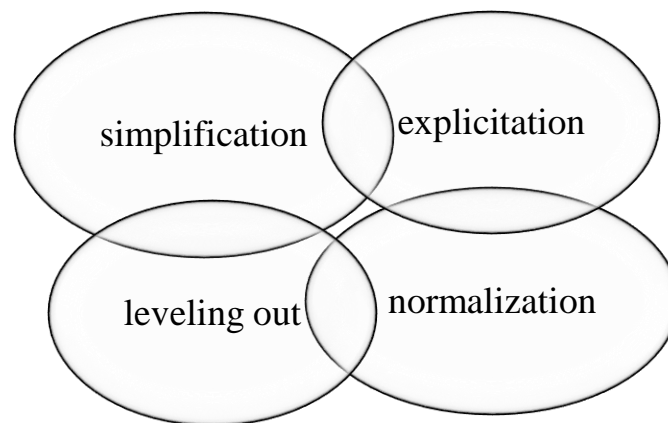


Table 1. Linguistic tendencies of corpus-oriented translation research

These phenomena became central concepts in corpus-oriented translation research.

Another significant contribution to corpus-based translation research is the development of parallel and comparable corpora. Parallel corpora are divided into three groups. We can see these groups in the following diagram.

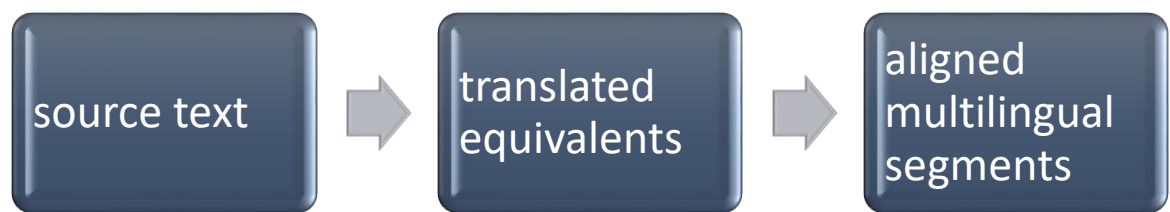


Diagram 1. Three groups of parallel corpora

These corpora allow researchers to examine equivalence relations, semantic transfer, lexical correspondence, and syntactic adaptation during the translation process.

Discussion

F. Zanettin emphasized that comparable corpora provide valuable insights into translator style, discourse preferences, and genre-specific translation tendencies [11; 41-p.]. Meanwhile, Mona Baker's Translational English Corpus (TEC) became one of the most influential corpora for studying variation and stylistic patterns in translated English.

S. Laviosa noted that translations often follow distinctive patterns that differ from non-translated texts, and corpora provide the tools necessary to explore and understand these patterns [4; 87-p.]. This data-driven perspective supports a more flexible and realistic understanding of translation as a dynamic and context-sensitive activity.

One of the biggest advantages of corpus-based research is its ability to handle vast amounts of data in a structured and replicable way. Researchers can trace how certain words, phrases, or structures are used in translated texts compared to original texts, offering insights into frequency, usage, and stylistic tendencies that may not be visible in smaller-scale studies.

One of the most important types of corpora in this kind of research is parallel corpora (reflecting source texts and their translations) and comparable corpora (representing original texts written in different languages or texts of the same genre written in different languages). Each type of corpus has distinct research goals: parallel corpora are used for studying specific translation strategies, while comparable corpora can be used for studying general trends and cross-linguistic similarities.

In addition, the development of a variety of software tools (including AntConc, WordSmith Tools, and Sketch Engine) has made the corpus analysis approach more accessible to scholars and students. These tools allow detailed searches and visualization of textual patterns so that both quantitative and qualitative analyses can be undertaken.

Conclusion

Based on the aforementioned research, we can identify different points of view. From our perspective, the usefulness of corpora goes beyond the scope of academic theory. As we have mentioned in other studies, corpora are a valuable tool for translation training, as they give students an opportunity to experience language as it is used in the real world. They provide exposure to stylistic norms, terminology, and the conventions of different genres. This enables students to make more confident and informed translation choices.

Professional translators also benefit from corpora, particularly when working within specific academic disciplines such as law, medicine, or technology. By consulting subject-specific corpora, translators can find reliable examples of technical terminology, idiomatic expressions, and context-appropriate phrasing, leading to more accurate and consistent translations.

Furthermore, corpora represent an important source of evidence about translation practice. They are used in modern translation research to provide a scientific basis for studying the behavior of translation professionals and as a tool for teaching and training translation scholars. Although corpora remain subject to certain limitations in modelling and representing statistical data that may obscure some important aspects of translation practice, they remain one of the most effective sources of information for understanding how translation works both in theory and in practice.

REFERENCES

1. Baker M. Corpus Linguistics and Translation Studies: Implications and Applications // Text and Technology: In Honour of John Sinclair / ed. M. Baker, G. Francis, E. Tognini-Bonelli. – Amsterdam: John Benjamins Publishing Company, 1993. – P. 233–250.
2. Blum-Kulka Sh. Shifts of Cohesion and Coherence in Translation // Interlingual and Intercultural Communication / ed. J. House, Sh. Blum-Kulka. – Tübingen: Gunter Narr Verlag, 1986. – P. 17–35.
3. Creswell J.W. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th ed. – Thousand Oaks, CA: SAGE Publications, 2014. – 273 p.
4. Laviosa S. Corpus-based Translation Studies: Theory, Findings, Applications. – Amsterdam; New York: Rodopi, 2002. – 152 p.
5. Laviosa S. Corpus-based Translation Studies: Theory, Findings, Applications. – Amsterdam; New York: Rodopi, 2002. – 152 p.
6. Laviosa S. Universals of Translation // The Routledge Encyclopedia of Translation Studies / ed. M. Baker. – London: Routledge, 1998. – P. 288–291.
7. Munday J. Introducing Translation Studies: Theories and Applications. 4th ed. – London: Routledge, 2016. – 376 p.
8. Olohan M. Introducing Corpora in Translation Studies. – London; New York: Routledge, 2004. – 218 p.
9. Bernardini S. Corpus-aided Translation Teaching // Using Corpora for Language Learning / ed. J. Sinclair. – Birmingham: University of Birmingham Press, 2004. – P. 97–115.
10. Wallmach K. Can a Translator Commit Suicide? Implications of Paul Ricoeur's Theory of Interpretation for Translator Training // South African Journal of African Languages. – 1997. – Vol. 17. – No. 4. – P. 119–126.
11. Zanettin F. Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies. – Manchester: St. Jerome Publishing, 2012. – 190 p.