AJLLSE GLOBAL RESEARCH NETWORK

AMERICAN Journal of Language, Literacy and Learning in STEM Education
Volume 3, Issue 12, 2025 ISSN (E): 2993-2769

Problems in Translating Uzbek Homonyms in Automated Translation Programs

Asilkhujaeva Diyorakhon Zikrillo qizi
Master's student, National University of Uzbekistan named after Mirzo Ulugbek

Abstract. *Homonymy is one of the widely observed linguistic phenomena in the Uzbek language and emerges in the process of language development. Homonyms, which are identical in form but different in meaning, demonstrate the multifaceted nature of the language. One of the key principles in identifying homonyms is context, as it plays a crucial role in determining the specific meaning in which a homonymous word is used. Therefore, the implementation of context-based translation algorithms is of great importance. This article focuses on the role of context in the translation of Uzbek homonyms by machine translation systems such as Google Translate, Yandex Translate, ChatGPT, and Tilmoch, and provides an analysis of the errors observed in these systems.*

Key words: *homonym, context, semantic relations, Google Translate, Yandex Translator, ChatGPT, Tilmoch.*

Introduction

In the context of the rapid development of modern information and communication technologies, machine translation systems have become an essential tool for facilitating communication between different languages. In particular, the effective performance of machine translation systems is directly dependent on the accurate analysis and interpretation of various linguistic phenomena, including polysemy and homonymy. In the Uzbek language, homonymous words are widely used, and identifying their meanings across different contexts remains a complex task for translation software. In many machine translation systems, the morphological and semantic characteristics of the Uzbek language are not sufficiently taken into account, which negatively affects translation quality. As a result, the accurate identification and interpretation of homonyms in translation systems has become a significant scientific and technical challenge at the intersection of modern linguistics and artificial intelligence.

This study examines the linguistic resources, methods, and approaches necessary to ensure the adequate translation of Uzbek homonyms in machine translation systems. In addition, the semantic, syntactic, and contextual aspects of homonym analysis are explored, along with the translation challenges they pose and possible ways to address these issues.

Relevance

The relevance of this study is determined by the fact that the incorrect translation of Uzbek homonyms can lead to misrepresentation of meaning and reduced translation quality, which, in turn, may cause significant errors in the translation of official and academic texts. A thorough investigation of this topic and the development of linguistically grounded solutions are important not only for machine translation systems but also for the field of general linguistics.

Purpose

The purpose of this study is to examine algorithms for identifying homonyms based on machine learning (ML) and artificial intelligence (AI), and to identify existing shortcomings in these systems through practical examples, providing recommendations for their resolution.

Main part

It is well known that homonyms are lexical units composed of identical phonemes but differing in meaning. According to the theory of expression, such units should be identical in form, considering both their pronunciation and spelling. M. S. Guricheva and B. A. Serebrennikov note that homonymy arises when new words, semantically unrelated, emerge from the division of word meanings. L. V. Malakhovskiy[1] defines homonyms as "words that are identical in pronunciation but differ in meaning, regardless of their spelling." Homonyms have been classified from various perspectives. The classifications developed in global linguistics, particularly in Russian linguistics, have influenced the study of homonyms in the Uzbek language[2]. At present, homonymy has become a central and pressing issue compared to other linguistic phenomena.

In languages such as English, Russian, Chinese, Japanese, and Italian, numerous studies in natural language processing (NLP) focus on the identification and automatic analysis of homonyms in machine translation, text processing, text comprehension, speech synthesis, and speech recognition. In the context of the Uzbek language, notable research on methods for identifying homonymous forms and translating them has been conducted by M. Abjalova[3]. Homonyms are among the most variable words in translation. For example, consider the English homonym *"bank."* In English, its first meaning corresponds to *"bank institution"* in Uzbek, while the second meaning refers to a *"container for holding liquid."* During translation, careful attention is required to preserve both meanings. In Uzbek, the word can be translated as *"bank,"* but context plays a decisive role. If we are talking about honey in a bank, the context indicates that the reference is not to a financial institution but to a *"container for holding liquid."*

The challenges of translating Uzbek sentences using machine translation primarily stem from the complex grammatical and semantic characteristics of the language. Machine translation systems such as Yandex Translate and Google Translate need to analyze the phraseological, morphological, and contextual aspects of the language more thoroughly to address these issues. Furthermore, translation algorithms should be further refined to account for the specific features of the Uzbek language. This, in turn, enhances the global understanding and usability of texts written in Uzbek.

During the translation of Uzbek texts, a number of linguistic challenges related to phonetics, morphology, syntax, semantics, and context emerge. The primary reason is that the Uzbek language has not yet been fully formalized for computational processing, and comprehensive linguistic support for translation systems has not been established. For example, the lack of a standardized transcription for Uzbek phonemes leads to deviations from normative pronunciation in voice-based translation applications. Additionally, misinterpretation of syntactic structures unique to Uzbek often results in syntactic errors, especially when complex sentence structures are translated using statistical machine translation systems. However, contemporary neural machine translation (NMT) systems, based on deep learning, have significantly reduced such issues, and most modern translation systems now attempt to avoid these errors.

Homonyms in the language can cause various problems in translation, including stylistic, morphological, and other issues. Recently, the presence of homonyms has led to numerous ambiguous

[1] Малаховский Л. В. Теория лексической и грамматической омонимии. – Л.: Наука. Ленинградское отделение, 1990. – 238 с.

[2] Abdurahmanova M. T. O'zbek va turk tillaridagi leksik omonimlarning lingvokognitiv xususiyatlari. filologiya fanlari doktori (DSc) dissertatsiyasi avtoreferati. – Toshkent, 2024. – 77 b.

[3] Abjalova M. Development techniques of analyzing homonyms in the languages that do not possess national corpus // "TURKLANG–2019: VII th International conference on computer processing of Turkic languages" Krim, 2019

and controversial situations in the translations produced by systems such as Google Translate. As homonymous forms continue to expand under the influence of extralinguistic factors, classifying words in context has become a valuable and essential resource in increasingly multilingual language pairs, as it accounts for both literal and figurative meanings. Nevertheless, automated translation cannot fully replace a human expert, since understanding the true contextual meaning requires a conceptual comprehension beyond a simple word-to-word correspondence. The table below presents selected homonyms from the *Explanatory Dictionary of the Uzbek Language*[4], chosen for experimental purposes, along with their translations and comparative analysis.

Table 1. Comparative Analysis of Machine Translation Systems

T/r	Sentence in the Uzbek Language	Google Translate Result	Yandex Translate Result	Tilmoch Result	Chat GPT Result	Commentary
1	**Milk**idan biroz tashlab tikish kerak.	It is necessary to sew with a little bit left.	It is necessary to sew a little from the tooth.	It needs to be stitched a little at the collar.	It needs to be sewn up a bit after being taken out of the milk.	Google Translate omitted the target word, whereas among the other systems, Tilmoch produced a relatively closer translation.
2	Ichimlikni **oyoq**qa solib uzatishdi.	They passed the drink on their feet.	The drink was put on his feet and passed.	They gave him the drink on his feet.	They handed the drink to me by putting it on the floor.	None of the translation systems was able to provide a fully correct translation.
3	**Otsiz**ni ko‘rsang, ehtiyot bo‘l.	If you see one without a horse, be careful.	Watch out if you see a horseman.	If you see a man without a horse, be careful.	Be careful if you see a horse.	In Google Translate and Tilmoch, only the primary meaning of the word was rendered, while the other two systems failed to take the suffix *“-siz”* into account.
4	**Rom**larni tuzatish uchun ko‘zdan kechirib chiqdik.	We have reviewed the roms to fix them.	We looked out to fix the ROMs.	We examined the frames to repair them.	We went through the documents to fix the mistakes.	Google and Yandex provided the word in an untranslated form, whereas ChatGPT produced a translation that was completely distant from the intended meaning. Nevertheless, Tilmoch generated a relatively accurate translation.
5	Tomlarda osilib turgan **sumalak**lar eriy boshladi.	The sumacs hanging from the roofs began to melt.	The sumac hanging on the roofs began to melt.	The icicles hanging on the roofs began to melt.	The sumalak hanging from the roofs began to melt.	Tilmoch provided an accurate translation, ChatGPT left the word untranslated, and the remaining two systems used a meaning related to spices in their translations.

[4] O‘zbek tilining izohli lug‘ati (birinchi jild). - Toshkent: O‘zbekiston milliy ensiklopediyasi, 2006. - 680 b.

6	U ko‘rpa-chasini dasturxon atrofiga **yozdi**.	He wrote the blanket around the table.	He wrote the blanket around the table.	He spread the blanket around the table.	He/she wrote the tablecloutharound the table.	Google, Yandex, and ChatGPT interpreted the word *“yozdi”* in the sense of “to write in a notebook,” whereas only Tilmoch produced the correct translation corresponding to “to spread out a quilt.”
7	Mushuk singlimni yuzini **tildi**.	The cat bit my sister’s face.	The cat wished my sister’s face.	The cat cut my sister’s face.	The cat scratched my sister’s face.	Only Tilmoch provided a correct translation.
8	Bobom ekkan daraxt **unib** chiqibdi.	The tree planted by my grandfather has sprouted.	The tree my grandfather planted sprouted.	The tree my grandfather planted had sprouted.	My grandfather's tree has sprouted.	All systems translated the word *“unmoq”* correctly; however, Google Translate produced a structurally more accurate sentence compared to the other systems.
9	Ukam boqayotgan **tana** ancha katta bo‘lib qolibdi.	The body that my brother is nursing has become much bigger.	The body that my brother was feeding became much larger.	My brother's body has grown much bigger.	The calf my younger brother is raising has grown quite big.	All translation systems rendered the word *“tana”* as “body,” referring to the human body, whereas ChatGPT correctly identified and translated its contextual meaning as “cow.”
10	Kelinning olib kelgan **sepi** juda ajoyib ekan.	The dowry brought by the bride is amazing.	The sepi brought by the bride turned out to be very cool.	The gift of the bride was very beautiful.	The dowry that the bride brought is truly magnificent.	Yandex Translate failed to translate the word *“sep”* and left it unchanged.

Discussion

The analysis indicates that among the four examined translation systems, Tilmoch provides the most optimal translations. The primary reason for this is its relatively richer linguistic resources tailored to the Uzbek language. By sufficiently enriching its linguistic database with Uzbek-language materials, the system is able to achieve more accurate and reliable translations. To attain comparable translation quality, other machine translation systems should likewise expand and enhance their linguistic support for the Uzbek language.

In conclusion, the main challenges in translating Uzbek homonyms into English include inadequate contextual understanding, the absence of direct lexical equivalents, and syntactic ambiguity. Addressing these issues requires the implementation of context-aware artificial intelligence systems, comprehensive lexical-semantic databases, and deep learning technologies. The development of translation systems that take into account the unique phonetic and semantic characteristics of the Uzbek language would not only improve translation quality but also make a significant contribution to the broader process of language digitization.

Results

Analyzing the translation of homonyms in Uzbek and English literary works allows for the formulation of several important conclusions. Undoubtedly, the translation of homonyms involves considerable difficulties due to the inherent “ambiguity” of this linguistic phenomenon. The concept of context plays a decisive role in determining accurate translations of homonyms, as it facilitates the selection of the most appropriate meaning. In the present study, identical homonymous sentences were translated using four translation systems—Google Translate, Yandex Translate, Tilmoch, and

ChatGPT. Based on the obtained results, Tilmoch was identified as the system providing the most optimal translations.

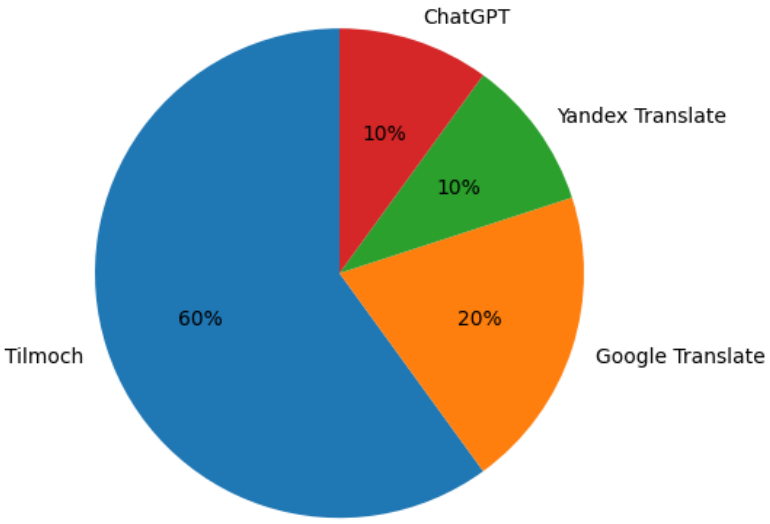

Figure 1. Distribution of Translators by Percentage

Based on the qualitative analysis of translation outputs, Tilmoch demonstrated the highest level of accuracy (approximately 55–60%), primarily due to its stronger contextual handling of Uzbek homonyms. Google Translate showed moderate performance (20–25%), while Yandex Translate (10–15%) and ChatGPT (around 10%) exhibited lower accuracy, often failing to account for contextual and morphological cues.

Conclusion

To achieve accuracy in machine translation, it is necessary to enrich an Uzbek homonym dictionary with semantic relations, that is, to construct semantic networks for each homonym. Such semantic networks enable translation systems to correctly disambiguate word meanings based on context. The semantic network of the Uzbek language can be developed using existing linguistic networks, such as WordNet and BabelNet. This approach significantly reduces semantic errors in automatic translation and enhances the overall accuracy of the system.

References

1. Малаховский Л. В. Теория лексической и грамматической омонимии. – Л.: Наука. Ленинградское отделение, 1990. – 238 с.
2. Abdurahmanova M. T. O‘zbek va turk tillaridagi leksik omonimlarning lingvokognitiv xususiyatlari. filologiya fanlari doktori (DSc) dissertatsiyasi avtoreferati. – Toshkent, 2024. – 77 b.
3. Abjalova M. Development techniques of analyzing homonyms in the languages that do not possess national corpus // “TURKLANG–2019: VII th International conference on computer processing of Turkic languages” Krim, 2019
4. O‘zbek tilining izohli lug‘ati (birinchi jild). - Toshkent: O‘zbekiston milliy ensiklopediyasi, 2006. - 680 b.