



Article

Addressing Lexical Ambiguity in English–Uzbek Neural Machine Translation

Abiyatova Muslima Maratovna¹, Qayimova Mohinur Saloxiddin qizi², Risboyeva Ozoda Bekzodova³

1,2,3. 3rd year student of SamSIFL Group 2302

*Correspondence: mabiyatova@gmail.com¹, mohinurqayimova1@gmail.com², ozodarisboyeva@gmail.com³

Abstract: One of the most difficult phenomena in the Machine Translation realm, and especially concerning English to Uzbek machine translation is polysemy — its meaning multiple in it. In this study, we explore the effect of polysemantic words on translation quality in terms of lexical ambiguity — contextual dependence and semantic disambiguation strategies. Employing a dataset encompassing twenty purposefully chosen sentences embedding words with high polysemy, the study explores the efficacy of modern Neural Machine Translation systems like Google Translate and DeepL. The results show that, as MT algorithms and neural network architectures have advanced, the same systematic deficiencies in resolving semantic ambiguity continue to exist for current systems. In particular, MT outputs are prone to statistical meanings over context-appropriate ones, resulting in errors specific domain interpretation, correct figurative language rendition, and homonym disambiguation. The study further proves that traditional methods including microglossarization and lexical choice codes, though partially useful, cannot simulate the human cognitive mechanism behind polysemy resolution as it depends on contextual correlation or associative networks. Enhancing translation accuracy and ensuring semantic correspondence between source and target texts also relies heavily upon lexical transformation methods—spanning specification, generalisation, differentiation, modulation and lexica compensation. In the end, working with polysemy serves not only to improve English-to-Uzbek machine-generated translation in terms of precision and naturalness, but also adds to the body of knowledge in a field that is concerned about complexity in linguistics more broadly, as well as cross-linguistic semantic relationship and computational perspective toward human-like language processing.

Keywords: Polysemy, machine translation (mt), lexical ambiguity, contextual disambiguation, lexical transformation, neural machine translation (NMT), semantic equivalence.

Citation: Maratovna A. M., Saloxiddin qizi Q. M., Bekzodova R. O. Addressing Lexical Ambiguity in English–Uzbek Neural Machine Translation. Central Asian Journal of Literature, Philosophy, and Culture 2026, 7(2), 265-269.

Received: 10th Jan 2026

Revised: 11th Feb 2026

Accepted: 19th Mar 2026

Published: 07th Apr 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

Introduction

There is always a human- an arbitrary signal (the voice sound)- that cover complex meanings. But the organic function of language bearing meaning over varied contexts is far too complex to be encapsulated by a few sentences [1]. The majority of problems in linguistic science are inextricably linked to the semasiological question, since it demanded a scientific investigation of communication by means of words. Out of these challenges, Polysemy is one of the main obstacles which needs to be addressed in making robust and intelligent Machine Translation systems [2]. Polysemy is the capacity for a word to have more than one unique definition. Highly developed polysemy, as a characteristic feature of language that indicates rich expressive resources [3], can be considered in both English and Uzbek languages. One word encompasses many concepts, making a vocabulary larger in terms of the expressive capability; however, it creates one “semantic bottleneck” for

computers that perform the logic. The chief difficulty is that a given word in a text may have multiple meanings, and the system has to determine which definition applies to a particular instance. Thanks to this disambiguation, the translation does not sound artificial, and loses the logical reasoning [4, 5] more often.

Literature review

The relationship between linguistic form and meaning has been a central topic of interest for philosophers and grammarians since classical antiquity, as in the work of Plato and Aristotle. The chapter you are reading is not based on a dry examination of roots, stems or affixes; with current linguistic science, it has become an advanced semasiology [6]. Of course, as John Lyons (1995) stresses: the natural function of language is to encode meaning but this becomes a complex task intrinsically because words have different sets of possible meanings even when there does not appear to be an obvious variation. While the "sameness" of meaning (synonymy) has long been a complicated problem, the "difference" of meaning — namely polysemy — one of the most insidious tasks in modern linguistics and machine translation. Theoretical work has identified two complementary processes of semantic development that dictate how a word's meaning grows over time: radiation and concatenation. A central or "primary" meaning serves as the anchor around which various secondary meanings may be arranged, radiating outwards like light rays in the process of radiation [7]. A classic example of this is the word "face". The central meaning, designating the human head's front aspect, is also the conceptual nucleus from which assorted metaphorical and functional extensions radiate, like face of a watch (soat sirti) or face of building (bino fasadi). On the contrary, Concatenation refers here to an incremental development of meaning where each subsequent one comes from the former one and tends to get quite far away from it [8] L. S. Barkhudarov (1975) offers even more nuance to this perspective, explaining that in any example of semantic change, it is not the meaning that is transferred, but that a particular word reassigns from one referent to another. This "transference" is explained by both historical changes in the economy and society as well as intralinguistic factors. Machine Translation systems face significant challenges without accounting for these patterns of evolution, often with little knowledge of whether a given secondary meaning is a direct extension or a semantic drift [9] and are incapable of not maintaining the "dialectical entity" of the word. In fact, the evolution of Machine Translation tells us about continuous efforts to overcome ambiguity resulting from polysemy. The early methods needed to spell out every definition of a word, demanding human involvement in choosing the right one. While this approach was inefficient and broke the logical flow of translation. Later developments, for instance microglossarization, have made this process more efficient by constricting potential meanings using the domain-specific context. An example of this could be the word "plan" which has different interpretations in biological and technical textual data [10].

Methodology

Lexical transformations while translating polysemantic words. For word-forming polysemantic words, various lexical transformations are used in the translation to maintain the meaning and consistency of the context.

Lexical Substitution. Lexical replacement means replacing one word with another one. They are often caused by differences between meaning structures of source language and target language words. The polysemantic word young is not the only one you would never translate as молодой: it does depend on combination of words — new potatoes, for example, would be молодой картофель. The equivalent is defined by the collocation of the word in use.

There are different types of deliberate lexical substitution as a translation technique:

a) Specification

This is replacing a word with a broad sense of meaning with one with a more narrow:
 1) Will you do the room? - Sen xonani tozalaysanmi?; 2) Get the papers on the way home.
 - Uyga ketish yo'lida gazetalar olib boring.

In this case, English polysemantic words have more general meanings than their Uzbek equivalents, and we can discern what exactly they mean only from the context.

b) Generalization

This consists in substituting a word that bears a restricted meaning for another that has an extended meaning: 1) People don't like to be stared at. → Этамалар статини ularning nishon asari.

The English verb stare refers to a more specific action than its Uzbek counterpart tikilib qaramoq, which can also mean to look or gaze at something.

c) Differentiation

This is a relatively rare phenomenon on this level, where one polysemantic word is substituted for another having the parallel meaning: 1) bamboo curtain → темир пәрда

Both bamboo and iron are hard in nature, both can be used as metaphor for political barriers (e.g., between two different ideological blocks).

d) Modulation

However, the rain outside rained.

The literal word for outside is tashqari but in Uzbek the translation that feels most natural is tashqarida or even ko'chada, depending on context. This accords with accepted usage and collocation norms.

Lexical Compensation. Lexical compensation means adding a new element in the target text intentionally to compensate for what has been lost in meaning or stylistic nuance from the source text.

Examples Showing Context Importance

a) *bright wine* → *bayramona vino*

b) *bright day* → *ajoyib kun*

c) *bright reply* → *yoqimli javob*

d) *a bright boy* → *aqli bola*

These examples clearly demonstrate that the meaning of polysemantic words is determined by context.

Results and Discussion

In this section, we present an empirical analysis on the impact of polysemantic words in English-to-Uzbek Machine Translation (MT) quality. This analysis arises from a small corpus: only twenty focused sentences with multivalent lexical items. These sentences were evaluated for translation by common Neural Machine Translation (NMT) systems, including Google Translate and DeepL. The findings show that even though modern MT systems are fluent and grammatical, they exhibit systematic limitations in resolving lexical ambiguity. Usually, the systems will revert to statistically dominant senses and not choose contextually appropriate interpretations [11].

Here are a few examples of the common failure cases in polysemantic disambiguation:

a) Domain Phone misinterpretation (The word "court")

Players are being trained on the court.

MT Output (Google Translate) O'yinchilar sudda mashq beryapti.

In this sentence, court means maydon or sports field. In fact the legal meaning (sud) is chosen wrongly by the MT system. It indicates that high-frequency links take precedence over contextual clues like the presence of word players within the system. This results in a pragmatically unacceptable translation [12].

b) High-frequency bias, in technical terms (the word “spring”)

The spring was broken, causing the clock to stop.

MT Output (Google Translate): It stopped, for spring was broken.

Spring – you know, the season (bahor), a water source (buloq) or even a mechanical part (prujina). The failure is also attributed to a lack of understanding about the technical context implied by the word clock, and instead a general meaning was chosen based on its highest usage in society [4]) Because of this, the translation becomes semantically wrong and illogical [13].

c) Idiomatic fake translation (the word “bright”)

She is the brightest student in the class.

MT Output (Google Translate): The brightest student in the class.

While yorqin is a direct translation of bright, it is used mainly in relation to physical light in Uzbek. Lest you misconstrue what I mean here, meaning intellectual ability and qi as in tuning it into zukko or aqli. This MT output shows that the system is not able to switch from literal decoding to understanding idiomatic meaning, therefore providing a stylistically unsuitable translation [14].

d) Homonymy and phonological ambiguity (the word “lead”)

In this old pipe, there is simply too much lead.

Since, there is a lot of lead in this old pipeline.

MT analyst predicts a potential pitfall where it confuses the noun lead (/led/, chemical element – qo'rg' oshin) with led verb / noun (/li:d/, leading – yetakchilik). Even in the context of a pipe, the system does not trigger the right semantic domain. This shows a limitation in processing homographs that have different pronunciation and meanings [15].

Conclusion

The analysis shows that polysemy still presents a major challenge to English → Uzbek MT, and although contemporary NMT systems can generate grammatic [sic] output for any given input, they also tend to misinterpret many polysemantic words due to their preference for the translation with the highest frequency (or more broadly speaking, higher metaphorical assessment) or due to literal translations. Such errors compromise both meaning and readability, as domain-specific context, figurative language and homonymy are often problematic. Lexical transformation strategies – including specification, generalization, differentiation, modulation, and lexical compensation – can reduce these problems by ensuring that the translated word matches the semantic field intended. More advanced context-specific disambiguation mechanisms may help make MT systems more similar to humans in handling polysemantic words. Hence, tackling polysemy is crucial not just for translation accuracy but also to maintaining the expressive wealth of language in English to Uzbek MT systems.

REFERENCES

- [1] C. N. Abdirashidovna and A. K. Nasimovna, "Features of translation of British articles of the internet media into the Uzbek language," *Eur. J. Innov. Nonformal Educ.*, vol. 1, no. 2, pp. 230–232, 2021.
- [2] C. Eshankulova, "Monosemy and polysemy of metaphorical terms in English and Uzbek," *Multidiscip. J. Sci. Technol.*, vol. 5, no. 12, pp. 1104–1108, 2025.
- [3] G. I. Hakimova, "A comparative linguistic analysis of lexical ambiguity and polysemy in English–Uzbek translation," *J. Int. Sci. Res.*, vol. 3, no. 8, pp. 1–12, 2026.

- [4] D. S. Q. Ibrohimova and N. M. Pazilova, "Semantic features of homonymy and polysemy in English and Uzbek," *Amer. J. Lang. Literacy Learn. STEM Educ.*, vol. 3, no. 4, pp. 279–282, 2025.
- [5] A. N. Kasimova, "Translation methods and techniques of publicistic materials," *Issue*, vol. 2, pp. 41–47, 2022.
- [6] A. N. Kasimova, "Must-know skills to translate newspaper articles," *Sci. Bull. Namangan State Univ.*, no. 3, pp. 281–287, 2023.
- [7] A. Kasimova and D. Nurullayeva, "Neologisms in modern English and their trends in word formation (based on texts in the media)," *Unversalnaya Indeksnyaya Biblioteka Evraziyskogo Zhurnala Sotsialnykh Nauk, Filosofii i Kultury*, vol. 3, no. 3, pp. 130–133, 2023.
- [8] N. Masharipova, "The functional features of polysemy in English and Uzbek languages," *Int. J. Innov. Technol. Soc. Sci.*, vol. 2, no. 23, pp. 49–52, 2020.
- [9] B. S. Pedersen, Ed., "Lexical ambiguity in machine translation," in *Recent Advances in Natural Language Processing*, John Benjamins Publishing Company, 2000.
- [10] A. Saodat, "A comparative study of lexical and grammatical polysemy in English and Uzbek: Linguistic and cultural insights," *SHOKH Library*, pp. 427–438, 2025.
- [11] N. Toreniyazov, "Comparative study of polysemy in English and Uzbek vocabulary," *Acad. Res. Mod. Sci.*, vol. 4, no. 55, pp. 49–51, 2025.
- [12] A. N. Kasimova, "Translation methods and techniques of publicistic materials," *Int. J. Word Art*, vol. 5, no. 3, 2022. (In Russian).
- [13] R. S. Madatovich, "The role of preschool education and family education in the raising of a healthy balanced generation," *For Teachers*, vol. 57, no. 4, pp. 520–523, 2024.
- [14] S. M. Ruzimurodov, "Ethnomadonian and ethnoesthetic aspects of the formation of a healthy lifestyle in society," *The American Journal of Social Science and Education Innovations*, vol. 3, no. 5, pp. 188–194, 2021.
- [15] A. A. Azizkulov and S. M. Ruzimurodov, "Gazali on the significance of intellect in human activity," *Paradigmata Poznani*, no. 2, pp. 43–46, 2014.