Next Scientists
European International Journal of Philological Sciences

TYPE Original Research
PAGE NO. 80-83
DOI 10.55640/eijps-05-10-17

Check for updates

Principles For Creating An Ontological Base For Mathematical Terms In English–Uzbek Machine Translation

Zokirova Khulkar Izzatillayevna

Senior teacher of Angren university, Uzbekistan

OPEN ACCESS

SUBMITED 30 August 2025
ACCEPTED 22 September 2025
PUBLISHED 28 October 2025
VOLUME Vol.05 Issue 10 2025

Abstract: Mathematical terminology often presents significant semantic ambiguity in machine translation (MT) systems, especially in language pairs with structural and conceptual asymmetry such as English and Uzbek. This paper discusses the linguistic and technological principles for creating an ontology-based lexical base aimed at enhancing the translation accuracy of mathematical terms. The proposed ontological model integrates linguistic analysis, conceptual structuring, and semantic relations to ensure consistent and contextually appropriate translation of polysemous terms. A bilingual corpus of English–Uzbek mathematical texts was used for term extraction, normalization, and semantic mapping within an ontology framework developed in Protégé using OWL/RDF representation. The research findings demonstrate that the ontology-based approach significantly improves semantic coherence and reduces contextual translation errors compared to conventional neural MT models. The proposed principles contribute to the development of a domain-specific linguistic base for English–Uzbek machine translation and can serve as a foundation for future multilingual ontological systems in scientific and technical fields.

Keywords: Machine translation, ontology, mathematical terminology, bilingual corpus, English–Uzbek, semantic disambiguation.

Introduction: Machine translation (MT) has become one of the most transformative areas in computational linguistics, especially with the advent of neural network architectures that have revolutionized cross-lingual communication (Koehn, 2020). However, despite significant advances, domain-specific translation—

particularly in mathematics—remains a challenging task. Mathematical language is highly abstract, symbolic, and context-dependent, which leads to substantial semantic ambiguity when translated automatically between typologically distinct languages such as English and Uzbek.

In English–Uzbek MT, mathematical terms often demonstrate polysemy and conceptual asymmetry. For instance, the English term “field” may refer to “maydon” in physics or “soha” in mathematics, depending on context. Conventional MT systems often fail to capture these nuances due to the lack of a structured semantic representation (Jurafsky & Martin, 2023). Ontological modeling, therefore, emerges as a promising approach to address these limitations by providing a machine-readable conceptual structure that defines entities, relationships, and domain-specific meanings (Gruber, 1993).

An ontology-based lexical base allows the MT system to interpret mathematical concepts within their correct semantic scope rather than relying solely on statistical co-occurrence patterns. Previous works such as WordNet and BabelNet have successfully demonstrated ontology-driven semantic interpretation in major languages (Miller, 1995; Navigli & Ponzetto, 2012), yet such resources for Uzbek remain underdeveloped. As a result, English–Uzbek MT systems lack the linguistic depth and conceptual structure needed for accurate translation of mathematical discourse.

This paper aims to formulate a set of linguistic and technological principles for constructing an ontological base of mathematical terminology applicable to English–Uzbek MT. The proposed framework integrates linguistic normalization, semantic role classification, and ontology structuring to enhance translation precision and terminological consistency. The study also discusses methodological insights on bilingual corpus alignment and the integration of ontology-based reasoning into neural MT pipelines. By bridging linguistic theory and computational modeling, this research contributes to the development of a scalable and semantically coherent ontology for English–Uzbek mathematical translation.

LITERATURE REVIEW

The concept of ontology has been widely explored in computational linguistics and artificial intelligence as a framework for representing structured knowledge. According to Gruber (1993), ontology is “a formal, explicit specification of a shared conceptualization,” enabling machines to interpret semantic relations among linguistic entities. Later, Noy and McGuinness (2001) refined this idea by outlining methodological steps for creating reusable and interoperable ontologies. Ontologies have since become a cornerstone of the Semantic Web, where structured knowledge representation allows intelligent systems to perform reasoning and disambiguation across languages and domains.

In machine translation, ontology-based approaches have been employed to overcome lexical and semantic ambiguities that cannot be resolved through statistical or neural methods alone (Navigli & Ponzetto, 2012). For example, EuroWordNet and TermNet projects demonstrated how interconnected concept networks can enhance cross-lingual semantic interpretation (Vossen, 1998). However, such resources rarely include lesser-studied languages such as Uzbek, which limits their utility in bilingual systems involving typologically distinct language pairs.

Mathematical terminology introduces additional complexity. Unlike ordinary vocabulary, mathematical terms are abstract, multi-domain, and often depend on symbolic or contextual interpretation (Nasukawa & Nagano, 2021). A single term can represent different meanings depending on its field of application—for instance, “ring,” “field,” or “root” have both mathematical and non-mathematical senses. In Uzbek, the challenge is compounded by the absence of structured terminological databases and standardized equivalents for many modern mathematical terms. Ontological modeling thus offers a promising solution by defining hierarchical concept relations, linking symbols with linguistic forms, and ensuring semantic precision across domains.

METHODOLOGY

The methodological framework of this research combines linguistic analysis and computational modeling to construct an ontological base for mathematical terms in English–Uzbek machine translation. The process follows a hybrid linguistic–technological approach that integrates corpus analysis, semantic classification, and ontology engineering.

A bilingual parallel corpus consisting of English and Uzbek mathematical textbooks, scholarly articles, and terminological dictionaries was developed as the primary data source. The corpus underwent linguistic preprocessing—tokenization, lemmatization, and part-of-speech tagging—to ensure consistency and accuracy during term extraction. Statistical and rule-based extraction techniques were used to identify domain-specific lexical units.

Each extracted term was subjected to semantic classification, where concepts were categorized based on their relations (e.g., subclass, attribute, function).

These semantic relations were then structured within an ontological hierarchy using Protégé software, encoded in OWL (Web Ontology Language) and RDF formats for machine interpretability. The ontology model includes three major levels:

1. Classes – representing broad mathematical domains (e.g., algebra, geometry, calculus);

2. Subclasses – specifying concepts within each domain (e.g., “group,” “vector,” “derivative”);

3. Instances – defining term-level entities linked to real textual examples from the corpus.

To integrate ontology into the MT pipeline, lexical–semantic disambiguation techniques were applied. These techniques allow the translation system to select the appropriate Uzbek equivalent based on contextual cues and relational mapping within the ontology. Neural MT models, such as Transformer and MarianNMT, were connected with the ontology layer using Python and SPARQL queries to perform dynamic term matching.

Key methodological steps:

• **Data source**: English–Uzbek mathematical textbooks and journal articles.

• **Preprocessing**: Tokenization, lemmatization, POS tagging.

• **Term extraction**: Hybrid statistical and rule-based approach.

• **Semantic modeling**: Class–subclass–instance hierarchy using Protégé.

• **Ontology representation**: RDF/OWL for machine readability.

• **Integration**: Ontology linked to neural MT models for improved semantic disambiguation.

This multi-layered methodology ensures that each term’s translation is context-aware, semantically grounded, and linguistically standardized—bridging the gap between linguistic theory and practical MT implementation.

RESULTS AND DISCUSSION

The ontology-based framework developed for English–Uzbek mathematical translation demonstrates a substantial improvement in contextual understanding and semantic accuracy. By linking mathematical terms to their conceptual representations, the model enables automatic disambiguation of polysemous terms. For instance, the English term “field”—ambiguous in ordinary usage—was correctly translated as “soha” in mathematical contexts and “maydon” in physical ones, illustrating the ontology’s ability to capture domain-sensitive meaning distinctions.

Quantitative evaluation was conducted using a test set of 2,000 bilingual sentence pairs derived from mathematical textbooks and research papers. The ontology-enhanced MT model, integrated into the MarianNMT framework, achieved a BLEU score improvement of 13.4% over the baseline neural MT system. Error analysis further revealed a 15% reduction in semantic inconsistency among symbol-dependent and multi-domain terms. Such improvements confirm that explicit conceptual modeling effectively supports linguistic coherence in translation.

Additionally, the ontology facilitated the generation of domain-specific bilingual glossaries, improving the accessibility and reusability of mathematical terminology. The integration of ontology-based semantic roles into translation pipelines improved lexical alignment between English and Uzbek concept nodes. This finding aligns with prior research emphasizing the role of ontology in enhancing MT interpretability and transparency (Buitelaar et al., 2018; Navigli, 2020). The developed system not only optimized term-level accuracy but also supported the standardization of Uzbek mathematical terminology—a key factor in expanding scientific communication across languages.

CONCLUSION

This study demonstrates that ontological modeling significantly enhances linguistic precision and semantic coherence in English–Uzbek machine translation of mathematical terms. By establishing structured conceptual relations among terms, the proposed ontology serves as an effective bridge between linguistic expression and domain-specific meaning. The research highlights the importance of integrating linguistic normalization, semantic role classification, and ontology structuring within the MT pipeline.

The findings confirm that ontology-driven MT systems reduce translation ambiguity, support terminological consistency, and promote the creation of standardized bilingual resources for low-resource languages such as Uzbek. Beyond improving translation quality, ontological modeling enhances system interpretability and provides a foundation for future integration with large language models (LLMs) and AI-based reasoning engines.

Future work will focus on the automatic expansion of the ontology, the integration of symbolic reasoning with neural translation architectures, and the development of multilingual ontologies for scientific domains beyond mathematics. The proposed approach thus represents a scalable and linguistically grounded step toward more accurate and semantically aware MT systems.

REFERENCES

1. Buitelaar, P., Cimiano, P., & Magnini, B. (2018). Ontology-Based Applications for Language Technologies. Springer.
2. Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199–220.
3. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.). Pearson.
4. Koehn, P. (2020). Neural Machine Translation. Cambridge University Press.
5. Miller, G. A. (1995). WordNet: A lexical database for English. Communications of the ACM, 38(11), 39–41.
6. Navigli, R. (2020). Natural language understanding: Instructions for (present and future) use. Communications of the ACM, 63(3), 46–53.
7. Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217–250.
8. Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory.