LINGUISTIC FEATURES OF REALIA IN CHINESE AND UZBEK LITERARY TEXTS: COMPUTER ANALYSIS

Dildora Umarova
Teacher at UzSIWL (Uzbek State World Languages University),
Andijan, Uzbekistan
Tel: +998946999628;
E-mail: dildoraumarova828@gmail.com

Abstract: This article comprehensively examines the linguistic features of national and cultural realia in Chinese and Uzbek literary texts using methods of computational linguistics and digital humanities. Specifically, it scientifically and practically substantiates algorithms for the automatic detection of cultural-domestic, historical, and geographical realia in translated texts, the creation of parallel corpora, and their comparison using semantic vector models.

Keywords: Computational linguistics, realia, national color, parallel corpus, semantic analysis, translation transformations, Chinese and Uzbek languages.

INTRODUCTION

Interlingual translation and comparative studies have always been of urgent importance in world linguistics. In particular, comparative study of concepts possessing national and cultural characteristics realia plays an essential role in the translation of literary texts in Chinese and Uzbek languages, which are typologically drastically different from each other. Given that traditional methods require substantial time and resources to process large volumes of texts, there is an increasing necessity today to make wide use of computational linguistics tools. By processing texts in digital format, applying statistical calculations and artificial intelligence algorithms, it is possible to evaluate the role and level of equivalence of realia in translation with high precision.

CHAPTER I. THEORETICAL FOUNDATIONS OF NATIONAL AND CULTURAL REALIA IN CHINESE AND UZBEK LITERARY TEXTS

1.1. Linguistic Nature of Realia and Their Classification

Realia refer to words or phrases inextricably connected with the lifestyle, history, material and spiritual culture, customs, and traditions of a specific nation, which require finding a specific equivalent or applying special translation transformations when translated into another language. While the Chinese language stands out for its foundation in hieroglyphic writing and richness in unique cultural codes, the Uzbek language possesses an agglutinative structure and is distinguished by its rich national-domestic color.

In linguistic research, it is advisable to study realia by dividing them into the following groups:

- Domestic realia: Household items, clothing, national dishes, and drinks (e.g., Chinese chao fan, qipao or Uzbek do‘ppi, lag‘mon, osh).
- Ethnographic and mythological realia: Religious beliefs, customs, mythological heroes, and holidays.
- Geographical and nature realia: Specific concepts related to landscape types.
- Socio-political and historical realia: State structure, historical titles, and positions.

1.2. Typological Differences Between Chinese and Uzbek Languages and Problems of Equivalence in Translation

© 2026 International Journal of Literature and Language Studies. All Rights Reserved.

The Chinese language (Sino-Tibetan language family) is monosyllabic, syllabic, and analytical, whereas the Uzbek language (Turkic language family) possesses agglutinative and synthetic features. This typological divergence creates serious difficulties in the translation process between the two languages. In particular, when a cultural realia expressed through a single hieroglyph or compound word in Chinese is translated into Uzbek, it often requires an expanded description or commentary.

CHAPTER II. METHODS OF COMPUTATIONAL LINGUISTICS FOR AUTOMATIC DETECTION AND COMPARATIVE ANALYSIS OF REALIA

2.1. Formation of Parallel Corpora and Pre-processing

The fundamental basis of computer analysis is a qualitative and structured digital corpus.

- Creation of parallel corpus: Literary works in Chinese and Uzbek (for example, Yu Hua's novel "To Live" (活着) and its Uzbek translations) are entered into the database in the form of parallel texts.
- Tokenization and segmentation: Chinese texts are divided into words using special segmentators (such as Jieba libraries), while morphological analysis and lemmatization (bringing words to their base form) are performed for the Uzbek language.

2.2. Statistics and Machine Learning Algorithms (TF-IDF and NER)

The following modern methods are used to automatically extract realia from text:

- Frequency and TF-IDF analysis: The frequency of lexemes that have few equivalents among texts or retain their form in translation is calculated.
- Named Entity Recognition (NER): Special dictionary databases and neural networks are used to identify and tag names, geographical names, and cultural objects.

2.3. Vector Models and Assessment of Semantic Similarity

Interlingual embeddings based on Word2Vec, FastText, or BERT are used to calculate how closely the contextual meanings of translated words match in both languages. If a word loses its original layer of meaning or is interpreted differently when translated, it is automatically recorded by the computer as a semantic shift (realia).

CHAPTER III. STATISTICAL-MATHEMATICAL VISUALIZATION OF TRANSLATION TRANSFORMATIONS

The Chinese and Uzbek realia identified as a result of computer analysis are grouped by translation methods and subjected to statistical calculations:

- Transcription and transliteration (phonetic assimilation).
- Calquing (literal translation method).
- Descriptive translation (explanations and commentaries revealing the meaning).
- Finding a functional analog (applying a national equivalent).

Software analysis results: Using the pandas, matplotlib, and seaborn libraries in the Python programming language, percentage ratios and application frequencies of these methods are displayed in clear diagrams. This makes it possible to show the achievements and losses made by translators in exact numbers.

CONCLUSION

Conducting a computer analysis of realia in Chinese and Uzbek literary texts serves to enrich translation theory with practical and empirical data. Methods of computational linguistics (creation of parallel corpora, statistical modeling, and semantic vector analysis) allow for an objective assessment of successful solutions in the translation process. In the future, creating special software modules and

© 2026 International Journal of Literature and Language Studies. All Rights Reserved.

automated translation analysis systems in this direction will be of immense scientific and practical significance.

REFERENCES

1. Explanatory Dictionary of the Uzbek Language. – Tashkent: "National Encyclopedia of Uzbekistan" State Scientific Publishing House, 2020.
2. Yu Hua. 活着 (To Live). Beijing, 1993. (Compared with Uzbek translations).
3. Vlahov S., Florin S. Непереводимое в переводе (The Untranslatable in Translation). — Moscow: International Relations, 1980.
4. Baranov A.N. Введение в прикладную лингвистику (Introduction to Applied Linguistics). — Moscow: URSS, 2001.
5. Ziyatova D. Representation of National and Cultural Realia in Chinese and Uzbek Literary Translation. Collection ofScientific Articles, 2024.

© 2026 International Journal of Literature and Language Studies. All Rights Reserved.