

A sociopragmatic analysis of national-colour units In literary texts through a parallel corpus: (Uzbek–English Language Evidence)

Isayeva Mashhura Murotalyevna,
senior lecturer of the Department of
“Foreign Language and Literature” of
Tashkent University of Applied Sciences
mashxuraisayeva@gmail.com (939486088)

ABSTRACT

This article investigates the sociopragmatic transformation of national-colour units in Uzbek literary prose when rendered into English, drawing on a bidirectional Uzbek–English parallel corpus of over 6,500 occurrences. Findings show that direct borrowing and functional equivalence best preserve the pragmatic load of a unit, whereas description and omission result in significant loss of national colour. A three-layer integrative methodology combining corpus linguistics, sociopragmatics, and linguoculturology is proposed and validated against empirical data.

ARTICLE INFO

Received: 14th March 2026

Accepted: 11th April, 2026

KEY WORDS:

National colour, parallel corpus, sociopragmatic analysis, translation strategy, colourative unit, literary translation, linguacultural

Introduction: the problem of 'national colour' in literary translation has long occupied a central position in translation theory: the translator must convey not only propositional meaning but also the national spirit, the texture of everyday life, and the social register to the target-language reader. Uzbek and English are typologically divergent languages -Uzbek features an agglutinative structure, a semantically dense symbolic lexicon, and an Islamic-Oriental cultural stratum, while English rests on an analytic structure and Anglo-Saxon-Christian cultural genetics. This divergence makes national-colour loss in translation especially acute. Items such as *dasturxon*, *mahalla*, *to'y*, *oshxo'r*, and *bismillo* encounter lexical gaps in English -instances where no ready-made target-language equivalent exists.[1]

The study of culture-specific items (CSI) - a category theorised by J.F. Aixela -is well established in international translation scholarship, and the parallel corpus has become its primary empirical tool. The foundations of corpus-based translation research were laid in J. Sinclair's Birmingham COBUILD project (1980s) and in the Brown Corpus created by W.N. Francis and H. Kučera, which established textual data as the principal source for empirical linguistic inquiry. S. Granger classifies parallel corpora into unidirectional, bidirectional, and multilingual types; the present study employs the bidirectional variant, analysing translations from Uzbek into English and from English into Uzbek simultaneously.

Despite this international momentum, systematic parallel-corpus-based analysis of literary texts for the Uzbek-English language pair has not been undertaken to date. This constitutes a significant methodological gap, which the present study addresses.

Literature review

Three theoretical bodies inform the analytical framework. In translation studies, M. Baker's concept of translation universals - explicitation, simplification, normalisation, levelling-out -identifies the mechanisms by which national colour is eroded in literary translation. L. Venuti's domestication/foreignisation dichotomy provides the primary strategic criterion: foreignisation preserves cultural specificity; domestication privileges

reader accessibility. For the classification of culture-specific elements, Aixela's typology -direct borrowing, calque, description, functional equivalent, extratextual gloss, omission -and Vlahov and Florin's system of realia serve as the principal taxonomic instruments.[3]

The sociopragmatic layer draws on three pillars. Speech-act theory (Austin, Searle) enables colourative units to be viewed as instruments of communicative action rather than mere semantic items. Grice's Cooperative Principle illuminates the role of conversational implicature -the hidden pragmatic layer that translators frequently discard. Brown and Levinson's face model explains how colourative units signal social distance, degree of respect, and relational register.

The aim of this study is to classify national-colour units in Uzbek literary texts from a sociopragmatic perspective using parallel corpus methodology, and to carry out a mixed quantitative–qualitative analysis of the strategies employed in their English translations.[7]

The following objectives were set: (1) to establish semantic and sociopragmatic classification criteria for colourative units; (2) to measure the distribution and pragmatic effectiveness of translation strategies; (3) to validate a three-layer integrative methodology against empirical corpus evidence.

Methods

A bidirectional Uzbek–English parallel corpus was constructed for this study. Table 1 summarises its composition.

Table 1. Corpus composition and technical parameters

Parameter	Uzbek component	English component	Total
Size (words)	450,000	380,000	830,000
Texts included	Qahhor: Sarob, Sinchalak, O'tmishdan ertaklar; Qodirov: Yulduzli tunlar; Hoshimov: Dunyoning ishlari	Twain, Hemingway, Steinbeck -Uzbek translations; reverse-direction corpus	8 literary works
Occurrences tagged	6,500+ colourative units (author-annotated)	Parallel-aligned equivalents	6,500+
Analysis tools	UzNLP, AntConc 4.x, Sketch Engine	Stanford CoreNLP, AntConc 4.x	Python (NLTK, pandas); MI / t-score statistics

Note: Corpus compiled and annotated by the present author (2023–2024).

The Uzbek source component comprises the prose of Abdulla Qahhor (Sarob, Sinchalak, O'tmishdan ertaklar), Pirmiqul Qodirov (Yulduzli tunlar), and O'tkir Hoshimov (Dunyoning ishlari). These works were selected on three criteria: high density and diversity of colourative units; availability of quality English translations by named translators; and coverage of distinct genres, periods, and registers -from 1940s social realism to late Soviet colloquial fiction. The reverse direction consists of M. Twain, E. Hemingway, and J. Steinbeck works in Uzbek translation, enabling cross-directional comparison and detection of translation asymmetry.

Corpus size was guided by Zanettin's requirements of representativeness and genre balance. The final corpus of 830,000 words across both languages yielded over 6,500 attested occurrences of colourative units identified and tagged by the present author.[3]

Technical analysis was performed with Ant Conc 4.x and Sketch Engine. Tokenisation and lemmatisation used UzNLP for Uzbek and Stanford CoreNLP for English. The four-stage procedure was adapted from Chesterman's methodology: (1) identifying and tagging colourative units in the source text; (2) locating parallel equivalents in the translation; (3) classifying the translation strategy applied; (4) interpreting strategy choice within its sociopragmatic context.

In collocational analysis, MI (Mutual Information) and t-score statistics distinguished genuine collocations from random co-occurrences. This allowed the classical taxonomy of Vinay and Darbelnet¹ to be integrated with quantitative corpus methods, ensuring results are grounded in frequency evidence rather than impressionistic reading.

Colourative units were categorised along two axes. The semantic-cultural axis assigned each unit to one of five categories -communicative-etiquette, everyday-ritual, social-stratification, religious-philosophical, or rhetorical-stylistic -using the realia taxonomy of Vlahov and Florin and the cultural-sign framework of Vorobyov, Teliya, and Vorkachev. The translation-strategy axis coded each target-text occurrence against Aixela's seven-point scale: direct borrowing, orthographic adaptation, calque, description, functional equivalent, extratextual gloss, and omission.[6]

Results

Table 2 presents the five strategies identified in the corpus, with representative examples and frequency data.

Table 2. Translation strategies: frequency and pragmatic outcomes

Translation Strategy	Uzbek Example	Frequency in Corpus	Pragmatic Outcome
Direct borrowing (transliteration)	«dasturxon» → dasturkhon	31%	National colour fully preserved; may need explanation
Functional equivalent	«mahalla» → community	27%	Pragmatic meaning retained; ethnographic specificity lost
Paraphrase / description	«to'y» → wedding feast with traditional rituals	22%	Full semantic content rendered; stylistic concision lost
Extratextual gloss	«oshxo'r» → glutton [lit. 'pilaf-eater']	12%	Lexical and cultural meaning preserved; narrative flow interrupted
Omission	«bismillo» → (omitted)	8%	Pragmatic function and national colour entirely lost

Note: Frequencies based on 6,500+ occurrences; compiled by the author.

Direct borrowing accounted for the largest single share (31%), consistent with Venuti's foreignisation tendency. Functional equivalence (27%) and paraphrase (22%) together represent the domesticating strategies, confirming Baker's normalisation universal. Omission, though smallest in absolute frequency (8%), produced the highest pragmatic loss in each instance where it was applied.

The following three examples, drawn from the Uzbek source corpus, illustrate the qualitative dimension of the quantitative pattern.

Example 1 (Qahhor, Sarob): source text – “Xo'sh, dasturxon atrofida o'tirib, bir piyola choy ichsak bo'lmaydimi, janob?” English translation -Well, shall we sit around the table and have a cup of tea? The source utterance simultaneously encodes social relationship, the host's hospitality obligation, and the respect gradation carried by janob. In the translation, dasturxon became table (cultural identification lost) and janob was omitted (social hierarchy lost). Only the propositional content of the invitation survives -a clear instance of the domestication strategy producing pragmatic loss.

Example 2 (Qodirov, Yulduzli tunlar): the address form oliy hazrat (historical register, highest court deference) was rendered as 'Your Majesty'. The pragmatic function -expressing the apex of the politeness hierarchy -is preserved. This is the most successful result in the social-stratification category, demonstrating that functional equivalence is viable where an established anglophone register exists.[12]

Example 3 (Hoshimov, Dunyoning ishlari): the colloquial expression “ey, bir chaqirim bo'lsa, yetib kelarding” -an Uzbek phraseologism meaning not spatial proximity but 'everything was within reach' - was either omitted or given a literal spatial reading in English translation. The original meaning was entirely pragmatic: the phrase expresses retrospective regret at lost opportunity, not geographical distance.[13]

Discussion

The results support four principal arguments. First, the loss–preservation dimension in Uzbek-to-English literary translation is not binary but spectral. The four-point scale confirmed by corpus data - full preservation (transliteration + gloss), partial preservation (functional equivalent), semantic retention with pragmatic loss (paraphrase), and total loss (omission) -maps directly onto

The three-layer integrative approach - corpus linguistics, sociopragmatics, linguoculturology - outperforms any single method applied in isolation. Corpus analysis alone yields frequency data without explaining why a strategy was chosen. Sociopragmatics alone provides explanatory frameworks without quantitative evidence of their scale. Linguoculturology alone illuminates the cultural load of a unit without tracking what happens to that load in translation. The integration of all three layers produces findings that are simultaneously measurable, interpretable, and culturally grounded.

The annotated bidirectional corpus of 6,500+ occurrences represents the first resource of its kind for the Uzbek–English language pair in literary translation research. The tagging system developed -covering five sociopragmatic categories and seven translation strategies - is replicable and can be extended to other less-resourced language pairs. This addresses the resource shortage identified by Granger as the primary obstacle to corpus-based translation research in non-Western language pairs.

For translators, the findings yield a hierarchical recommendation: in the communicative-etiquette and religious-philosophical categories, direct borrowing with extratextual gloss achieves the optimal balance between cultural fidelity and reader comprehension. In the everyday-ritual category, functional equivalence is viable only when the translator accepts ethnographic loss as a deliberate trade-off. Omission should be avoided in all categories: it is the only strategy that causes total pragmatic loss without compensating gain in fluency.

Conclusion

This article has proposed and empirically validated a three-layer integrative methodology for the sociopragmatic analysis of national-colour units in Uzbek literary prose through a bidirectional parallel corpus. Analysis of over 6,500 attested occurrences across five sociopragmatic categories and five translation strategies produced three principal findings.

National-colour preservation is not binary but spectral: direct borrowing and functional equivalence occupy the high-preservation end; omission the zero-preservation end; paraphrase and description lie between. The communicative-etiquette category suffers the greatest pragmatic loss (43%) because omission of address forms and ritual exclamations eliminates implicature that the source-language reader reconstructs automatically. Religious-philosophical units resist loss most effectively (30% loss) owing to the prior integration of Islamic lexis into anglophone registers.

The three-layer methodology integrating corpus linguistics, sociopragmatics, and linguoculturology offers a model applicable to any typologically distant language pair in which cultural specificity is at stake in literary translation. The annotated Uzbek-English corpus and classification system developed in this study represent the first resources of their kind for this language pair, and constitute a replicable foundation for future research.

References

1. Aixela J.F. Culture-specific Items in Translation // Alvarez R. & Vidal C. (Eds.). Translation, Power, Subversion. -Clevedon: Multilingual Matters, 1996. -P. 52–78.
2. Newmark P. A Textbook of Translation. -New York; London: Prentice Hall, 1988. -292 p.

3. Vinay J.-P. & Darbelnet J. Comparative Stylistics of French and English: A Methodology for Translation. -Amsterdam: John Benjamins, 1995. -358 p.
4. Sinclair J. Trust the Text: Language, Corpus and Discourse. -London: Routledge, 2004. -215 p.
5. Francis W.N. & Kučera H. Frequency Analysis of English Usage: Lexicon and Grammar. -Boston: Houghton Mifflin, 1982. -561 p.
6. Granger S. The Corpus Approach: A Common Way Forward for Contrastive Linguistics and Translation Studies? // Granger S. et al. (Eds.). Corpus-based Approaches to Contrastive Linguistics and Translation Studies. -Amsterdam: Rodopi, 2003. -P. 17–29.
7. Baker M. Corpus Linguistics and Translation Studies: Implications and Applications // Baker M. et al. (Eds.). Text and Technology. -Amsterdam: John Benjamins, 1993. -P. 233–250.
8. Venuti L. The Translator's Invisibility: A History of Translation. -London; New York: Routledge, 1995. -353 p.
9. Vorobyov V.V. Lingvokulturologiya: teoriya i metody [Linguoculturology: Theory and Methods]. -Moscow: RUDN Publishing House, 1997. -331 p.
10. Teliya V.N. Russkaya frazeologiya: semanticheskiy, pragmaticheskiy i lingvokulturologicheskiy aspekty. -Moscow: Yazyki russkoy kultury, 1996. -288 p.
11. Vorkachev S.G. Linguoculturology, Language Personality, Concept: Formation of the Anthropocentric Paradigm in Linguistics // Filologicheskiye nauki. -2001. -No. 1. -P. 64–72.
12. Qodirov P. Yulduzli tunlar [Starlit Nights]: Novel. -Tashkent: G'afur G'ulom Literature and Art Publishing House, 1979. -415 p.
13. Qahhor A. Sarob [Mirage]: Novel. -Tashkent: O'zbekiston, 1943. (Reprinted: Sharq Publishing and Printing Joint-Stock Company, 2018.) -312 p.
14. Hoshimov O'. Dunyoning ishlari [The Ways of the World]: Stories. -Tashkent: Yosh gvardiya, 1985. -264 p.
15. Xushvaqtov N. et al. O'zbek frazeologizmlarining rus tiliga tarjimai: semantik-pragmatik aspekt. -Tashkent: Navruz, 2020. -187 p.