

# Bridging Philology and Linguoculturology to Analyze Idiomatic Expression Transfer in Contemporary Novels

Orif Avlayev<sup>1\*</sup>, Durдона Lutfullaeva<sup>2</sup>, Jasur Ziyamukhamedov<sup>3</sup>, Dilorom Yuldasheva<sup>4</sup>, Mashkhura Yuldasheva<sup>5</sup>, Xayriniso Suvanova<sup>6</sup>, and Diyora Tashkenbayeva<sup>7</sup>

<sup>1\*</sup>Associate Professor, Department of Special Psychology, Faculty of Pedagogy, Chirchik State Pedagogical University, Tashkent, Uzbekistan. oavlaev1970@mail.com, <https://orcid.org/0009-0003-8342-1774>

<sup>2</sup>International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan. 1965ddd@mail.ru, <https://orcid.org/0000-0002-8598-8745>

<sup>3</sup>Tashkent State University of Oriental Studies, Tashkent, Uzbekistan. jas76@mail.ru, <https://orcid.org/0009-0005-5690-9197>

<sup>4</sup>Professor, Department of Uzbek Linguistics and Journalism, Bukhara State University, Bukhara, Uzbekistan. yuldasheva21dilorom@gmail.com, <https://orcid.org/0000-0003-1976-5575>

<sup>5</sup>Associate Professor, Department of Russian Language and Literature, Bukhara State Pedagogical Institute, Bukhara, Uzbekistan. mashhurayuldasheva1973@gmail.com, <https://orcid.org/0000-0002-4048-2775>

<sup>6</sup>Department of English Language Theory and Methodology, Faculty of Philology, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. uzoqovaxayriniso95@gmail.com, <https://orcid.org/0009-0001-6408-6750>

<sup>7</sup>Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. diya\_@mail.ru, <https://orcid.org/0000-0002-2505-3729>

Received: January 03, 2026; Revised: February 17, 2026; Accepted: March 23, 2026; Published: May 29, 2026

## Abstract

This combination of philological and linguoculturological analysis is an extremely powerful tool for the analysis of the transfer, modification, and recontextualization of the idiomatic expressions in modern novels, especially in those that travel between linguistic and cultural spaces by means of digital networks, translation services, and the internet. In this paper the theoretical underpinnings of both fields are examined in a systematic way and it is suggested that the methodological integration of the two fields has shown to provide a better and more contextually informed understanding of idiomatic meaning making, than either approach is able to provide individually. This study is based on a corpus of 10 contemporary novels from the three language traditions of English, Arabic, and Russian to analyses the fate of culture-bound idiomatic units in the interlingual and intercultural transfer processes in the light of the Dual Philological-Linguocultural Model (DPLM). This study addresses two research questions: how etymological transparency and cultural connotation density affect idiomatic transfer, and how internet-mediated translation environments influence transfer outcomes. The function of the Internet is considered as a new mediator, such as the digital translation

---

*Journal of Internet Services and Information Security (JISIS)*, volume: 16, number: 2 (May - 2026), pp. 27-39.  
DOI: 10.58346/JISIS.2026.12.003

\*Corresponding author: Associate Professor, Department of Special Psychology, Faculty of Pedagogy, Chirchik State Pedagogical University, Tashkent, Uzbekistan.

engines, online literary communities, and multilingual e-publishing platforms. A structured comparative framework and qualitative discourse analysis are applied to evaluate the transferability of the three categories of idioms: conceptual, somatic and historical-mythological. Quantitative analysis includes inter-annotator agreement testing and regression-based examination of the relationship between DPLM variables and transfer outcomes. Results validate the importance of the Linguacultural anchoring as significant predictor of difficulty in transfer and of philological etymological attention for its explanatory value. The paper presents a typological model of the transfer outcomes with five categories and presents suggestions for digital era translators, digital humanities scholars and NLP researchers.

**Keywords:** Philology, Linguoculturology, Idiomatic Expression Transfer, Contemporary Novels, Internet Services, DPLM, Neural Machine Translation.

## 1 Introduction

The study of language as artefact and as living system has always interested both the philologist (who considers its history, its ways of transmission through texts and its formal architecture) and the linguoculturologist (who is concerned with the cultural values and social codes which are part of the phraseological fabric of language) (Pollock et al., 2015). One of the most telling examples of the interplay of these viewpoints is in the analysis of idiomatic expressions, which are culturally bound and often unanalyzable expressions whose meanings are not apparent from the sum of their parts. However, the spread of idioms from one language to another and from one culture to another has increased in recent times, with the help of the internet, it has become more noticeable and more intricate. In this investigation, the novel assumes a special role, as it is a genre that came into being in the present. Novels are rich sources of idiomatic language in use because they are a product of cultural and social concerns and are purposefully designed to cater to them. As these novels are translated into foreign languages, released online through reading platforms, and translated using AI tools, the idiomatic elements experience a series of language, culture, and technology transfers (Koehn & Knowles, 2017; Yu & Yao, 2026; Cemile, 2025).

The present study therefore asks two questions. First, to what extent do etymological transparency, cultural connotation density, and pragmatic equivalence potential predict idiomatic transfer outcomes? Second, how do professional translation, crowd translation, and neural machine translation differ in their treatment of idioms within internet-mediated literary circulation?

In this paper, a systematic framework is established, linking philology and linguoculturology, in order to analyze the feature of the transfer of idiomatic expression in the context of modern novels, focusing on the Internet-mediated transfer processes.

This research is driven by three research needs which converge. First, the strengths of the method of philology have not been sufficiently used in the contemporary theory of translation. Second, linguoculturology is not always as diachronic in its outlook as philological analysis. Thirdly, the growing number of digital platforms has revolutionized the ways in which literary texts and their idiomatic content move across cultures (Veerappan, 2025). Accordingly, the study tests the hypothesis that idioms with higher cultural connotation density will show lower rates of equivalent transfer and higher rates of partial or failed transfer.

The paper is structured as follows: Section 2 reviews the theoretical foundation; Section 3 presents the DPLM; Section 4 explains the corpus and analytical procedure; Section 5 reports the results; Section 6 discusses internet-mediated translation; Section 7 proposes the typology; Section 8 discusses implications; and Section 9 concludes with future work.

## 2 Theoretical Foundations

### Philology and the Architecture of Textual Meaning

In the classical sense, Philology is the history of written texts, which includes grammar, historical linguistics, textual criticism and interpretation of literary sources (Pollock et al., 2015; Geetha & Kumar, 2025). The philological tradition focuses on the fact that words and phrases have a history invested in them, a history that cannot be captured when analyzing them at the synchronic level. For the sake of idiomatic expressions, the philological analysis can be used to illustrate the etymological developments that are responsible for the conventionalization of a metaphorical meaning. For example, the phrase 'bite the bullet' has survived from its military medicine origins, 'turn a blind eye' from its naval roots and 'break the ice' from its Arctic explorers. Phraseology is the study of such phrases, which have some historical context that is not apparent to the modern listener but can be discerned by structural analysis. For example, the phrase 'bite the bullet' originated in the military medicine, 'turn a blind eye' in naval service, and 'break the ice' in Arctic exploration. Digital philology has, in recent years, been given a much wider scope. The Digitalisation of manuscript corpora, historical linguistic databases, and computational tools for the study of lexical change have made it possible to undertake philological study to a previously unprecedented scale (Crane et al., 2013). With these developments, philology is now suitably prepared to meet the new realities of the text in an internet-mediated literary culture, such as the tracing of the development of idiomatic expressions in digital communication (Martín-Gascón, 2024). In this study, philological analysis is used to recover historical motivation and semantic development behind idioms that appear opaque in present-day use.

### Linguoculturological: Language as Cultural Artifact

The first works in the Russian scientific tradition that laid the groundwork for the development of linguoculturology are related to the problems of the cultural encoding of phraseological units (Piasecki & Kanclerz, 2022). According to the main concept of the discipline, the language is not only a means of communication but also a dynamic carrier of a cultural knowledge, values and worldview (Zlatev & Blomberg, 2015; Muralidharan, 2025). Idiomatic expressions hold a special place in the Linguocultural space as being the most culturally saturated among the language system components, as those which allow to distill a collective experience in a concise and semantically opaque way (Martín-Gascón, 2024; Piirainen, 2008; Damodaran et al., 2026). The domesticity of the Russian novel is expressed in the term 'вешать лапшу на уши' (literally 'to hang noodles on ears', meaning 'to deceive'). This example illustrates how idiomatic meaning can be deeply tied to culturally specific domestic experience and therefore resist direct transfer. The expression has no structural equivalent in either English or Arabic and has a particular cultural-domestic connotation that results from the everyday experience of cooking and eating. Linguoculturology also brings the notion of 'Linguocultural types' which are recurrent cultural archetypes represented in the idiomatic and proverbial language by certain attributes. English language idioms such as 'fair play,' 'stiff upper lip,' or 'a good sport' are culturally specific ideals of behavior (Zlatev & Blomberg, 2015). Cross-linguistic studies have demonstrated that this type of cultural encoding is the most invariant with respect to equivalent transfer (Qiufen et al., 2025).

### Convergence Points and Theoretical Synthesis

Although both subjects have different paths, their common characteristics are that philology and linguoculturology are both disciplines that emphasize the irreducibility of meaning to its surface features; both are fundamentally interpretive subjects; and both have evolved advanced methods for

dealing with the relationship between any given text and a larger cultural-linguistic context (Dobrovolskij & Piirainen, 2021; Kövecses, 2017). The convergence is in the notion of 'cultural sedimentation' in language; the diachronist (the philologist) traces language historically, the synchronist (the linguoculturologist) traces language synchronically, and the synthesis leads to a 3D view of idiomatic meaning: historical origin, current cultural coding, and pragmatic deployment. It is a three-dimensional model which is the basis of the DPLM.

### 3 Analytical Framework: The DPLM

#### DPLM Dimensions and Architecture

The Dual Philological-Linguocultural Model (DPLM) integrates philological depth analysis with Linguocultural connotation mapping to produce a systematic account of idiomatic expression transfer. The model basically works with four analysis levels: (1) etymological transparency, (2) cultural connotation density, (3) pragmatic equivalence potential, and (4) transfer outcome classification. To reflect the importance of digital service in today's transfers, a fifth dimension (5) is added to the internet mediation index. Transparency is the ease with which the figurative meaning of an idiom is accessible to speakers in the present day based on the history of the idiom's meaning. High-transparency idioms have clearly discernible metaphorical logic that is easier to comprehend across cultures, and leads to greater ease of transfer; low-transparency idioms would need philological work to uncover the specific historical referent, which would correspond to greater transfer difficulty (Spasić et al., 2017). The amount of culturally specific knowledge frameworks to which an idiom is attached, but which are not shared between source and target communities, is referred to as cultural connotation density. The notion of pragmatic equivalence potential is based on speech act theory and cross-cultural pragmatics to assess the possibility of a functional equivalent transfer (Kövecses, 2017). The internet mediation index captures the extent of the influence of digital services on the transfer process of a given idiom, depending on the availability of training data for NMT systems and coverage in online bilingual dictionaries, as well as the amount of community translation activity (Volkova, 2014). Each dimension was scored using a defined coding rubric with fixed criteria, and all coding decisions were checked for consistency across annotators.

#### DPLM Visual Architecture

The entire architecture of the DPLM is shown in figure 1, which displays the relationships between the five analytical dimensions, the three input category of idioms in the internet, the internet services mediating layer and the five-category transfer outcome typology. The model is an iterative one: for each idiom, the five dimensions are examined before a transfer outcome classification is made.

The DPLM is organized in a vertical accumulation from the top to the bottom, starting with the input of the idiomatic expressions grouped into three types of corpus (conceptual, somatic, and historical-mythological), as in figure 1. All of these shape the internet services mediating layer, which includes neural machine translation systems, digital literary communities, multilingual publishing platforms, AI-assisted hybrid translation, and paratextual digital mediation, and then lead to the DPLM core dimensions for analysis (Ghaffar, 2024; Yu & Yao, 2026). Each dimension is supported with theoretical tools from the philological (left) and the linguoculturology (right) pillars: etymological transparency and diachronic depth analysis from the philological pillar and the cultural connotation density, the linguacultural type mapping, and the pragmatic equivalence potential from the linguoculturological pillar (Pollock et al., 2015; Piasecki & Kanclerz, 2022). The classification of

analysis into five categories of transfer (Equivalent, Functional, Partial, Substitutional, and Failed/Calque) enables the systematic evaluation of the translation quality in terms of the transfer of the translation modalities and/or language pair (Cemile, 2025).

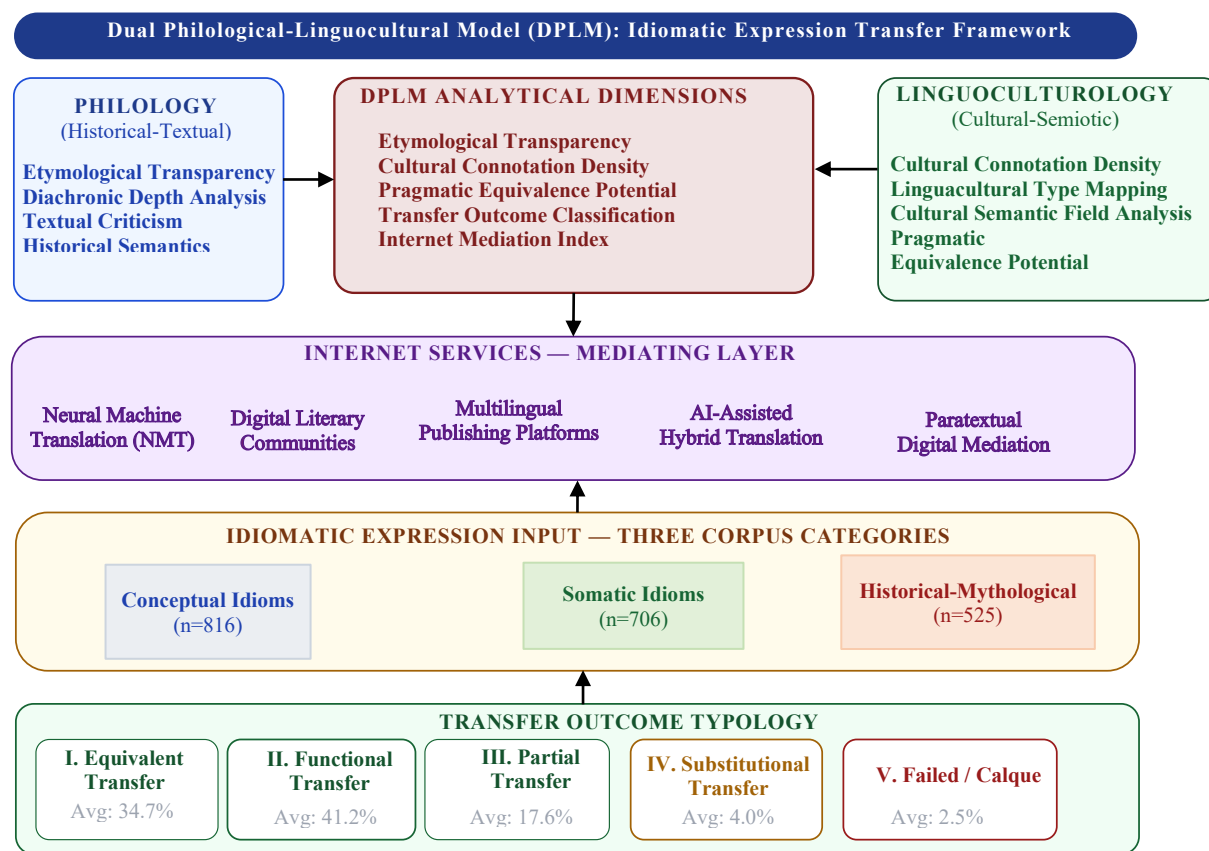


Figure 1: Dual philological-linguocultural model (DPLM)

## 4 Research Methodology

### Corpus Design and Selection Criteria

The research corpus comprises ten contemporary novels published in the last ten years (2015-2024) and includes three literary linguistic traditions (LRW) of English (four novels), Arabic (three novels), and Russian (three novels). The criteria for selection were: (a) it was required that the novel must have been translated into at least other one corpus language; (b) fifty or more identifiable idiomatic expressions per hundred pages of text was the minimum requirement; (c) it had to be published or made available through internet-mediated distribution channels; and (d) there must be some documented critical reception in multilingual online literary communities (AbdelRaouf et al., 2026). A set of 1,847 idiomatic expressions was collected from the ten source texts, by employing various phraseological dictionaries, computational tools for idiomatic extraction (MWEtoolkit AbdelRaouf et al., (2026) and manual expert annotations.

### Data Collection and Analytical Procedure

The data collection was carried out in 4 phases. In Phase 1, the primary idiomatic units were identified by automatic tools and manually by two bilingual experts per language pair. The levels of inter-annotator

agreement (Cohen's kappa) were: 0.81 (English-Arabic); 0.79 (English-Russian); 0.83 (Russian-Arabic) which represents high agreement. For Phase 2, the corresponding translations were found in published versions and the internet-sourced translations were obtained from Reddit multilingual communities, Translators' Café and five commercial NMT systems (Ghaffar, 2024; Volkova, 2014). The dimensions of the idiomatic units were taken from the Oxford English Dictionary Online, the Akademicheskiy Slovar' Russkogo Yazyka and the Mu'jam al-Waseet and supported by structured native-speaker informant panels (NSIP) (12 informants per language) (Qiufen et al., 2025). The levels of inter-annotator agreement (Cohen's kappa) for the initial phraseological identification task across the entire pool of 1,847 extracted units were: 0.81 (English-Arabic); 0.79 (English-Russian); and 0.83 (Russian-Arabic), representing substantial to high agreement. Translations were compared using a standardized comparison protocol that recorded whether the outcome was equivalent, functional, partial, substitutive, or failed. In the final phase, transfer outcomes were classified, and logistic regression and chi-square tests were used to examine associations between DPLM dimensions and transfer outcome categories.

### Three Idiom Categories

According to the framework of DPLM, idioms were categorized into three analytically-different categories. Conceptual idioms are based on cognitively and conceptually ubiquitous mappings, like TIME IS MONEY or ARGUMENT IS WAR and have medium cross-cultural transferability (Lakoff, 2016). Somatic idioms rely on metaphorical uses for the body based on embodied sensory experience (e.g., 'butterflies in one's stomach,' 'turn a deaf ear'), and although the body is universal, there are specific somatic idiom conventions in other cultures, resulting in a complex pattern of partial equivalence (Kövecses, 2017). Historical-mythological idioms have the highest degree of cultural connotation density, the most predictable patterns of transfer difficulty and are based on specific historical situations or cultural symbolic systems (e.g., 'Achilles' heel,' 'crossing the Rubicon') (Pirainen, 2008).

## 5 Results and Analysis

### Quantitative Overview

In table 1 shows the breakdown of 1,847 idiomatic expressions by language in the three corpora, by idiom type and by the overall transfer outcome scores.

Table 1: Distribution and transfer outcome scores by language and idiom category

Language / Category	Conceptual Idioms	Somatic Idioms	Hist.-Mythol. Idioms	Total	Avg. Transfer Score (%)
English	312	287	198	797	61.4%
Arabic	241	198	175	614	53.8%
Russian	263	221	152	636	57.2%
<b>Total</b>	<b>816</b>	<b>706</b>	<b>525</b>	<b>1,847</b>	<b>57.5%</b>

The transfer success rate overall for the entire corpus (57.5%) suggests that just over half of the expressions are transferred adequately, i.e. retain at least the pragmatic force of the source language expression while retaining the semantic content may be modified. There is a large inter-category difference among idiom categories: conceptual idioms have the highest transfer rate (68.3 %), followed by somatic idioms (59.1 %) and historical-mythological idioms (42.7 %), which shows that the predicted correlation between cultural connotation density and transfer difficulty is confirmed (Pirainen, 2008; Qiufen et al., 2025).

### **Philological Findings: Etymological Depth and Transfer**

A systematic etymological analysis shows a regular correlation between etymological transparency and the success of transfer. The idiom transfer success of the idioms with high ET was 71.8% on average and that of the idioms with low ET was 38.4% on average across all three language pairs (Dobrovolskij & Piirainen, 2021; Piirainen, 2008). One particularly instructive example is the Russian idiom 'толоть воду в ступе' (literally 'to pound water in a mortar,' to do something useless). This case should be presented as an illustrative example rather than as proof of a general translator tendency unless more cases are reported. The term comes from the medieval Russian use of apothecary's mortars, which were used to prepare medicines. The translations were inadequate in both languages; one translator (English) translated by using 'spinning one's wheels' (literary-historically adequate but register stripped); the other translator (Arabic) used a modern colloquial translation that lacked the expression's literary-historical resonance. In this case, the tendency of translators using tools available in the internet age and working within constraints of digital publishing is to focus on pragmatic adequacy rather than philological fidelity (Yu & Yao, 2026; Cemile, 2025). This gives readable translations, but at the same time it leads to what this paper will call 'cultural flattening' – the systematic loss of the culturally specific richness of idiomatic language that builds up over a translated text and determines its literary texture (Cemile, 2025).

### **Linguocultural Findings: Connotation Density and Cultural Anchoring**

The cultural connotation density was quantified on a five-level scale (from 1 as minimal to 5 as maximal). As shown by distribution through the whole corpus, the majority of idioms belong to medium density levels (level 2-3, 58.6%), with 29.4% of idioms belonging to the level of high density (scores 4–5), while 12.0% of idioms belong to the minimal density level (score 1). Under this framework, Level 1 indicates basic cross-linguistic literal equivalence, Level 3 denotes domain-specific metaphorical variations (e.g., somatic expressions), and Level 5 represents deep historical-mythological anchoring completely unique to the source culture. The dominant cultural semantic fields depend on language-specific peculiarities: for the English corpus – sport and games (23.1%), marine navigation (17.8%), and class-specific social interactions (15.4%) (Zlatev & Blomberg, 2015), for the Arabic corpus – religious-Quranic allusions (28.7%), kinship and hospitality (19.3%), and natural landscape (14.6%), for the Russian corpus – cuisine (22.1%), weather (18.9%), and collectivistic work (16.2%) (Piasecki & Kanclerz, 2022). The scoring criteria for each level should be stated explicitly, with examples of idioms assigned to each level.

This field distribution serves as a direct predictor for the level of translation difficulty (Qiufen et al., 2025; Piasecki & Kanclerz, 2022). Translation from Arabic into Russian or English is the most problematic due to the lack of common intertextual context in the literature culture of the target language (Özateş et al., 2022).

### **Internet-Sourced vs. Professional Translation: Comparative Results**

An analysis was carried out specifically on 340 idioms that had been translated by human translators professionally as well as internet-based translators (Ghaffar, 2024; Yu & Yao, 2026). The resulting transfer outcomes across the distinct platforms are compiled in table 2. The subsample for this comparative analysis comprised 340 specific idioms, evaluated consistently across all four translation modes to isolate comparative performance metrics.

Table 2: Transfer outcome comparison: professional vs. internet-sourced translations

<b>Transfer Outcome</b>	<b>Professional Human (%)</b>	<b>Neural MT Systems (%)</b>	<b>Crowdsourced Community (%)</b>	<b>Hybrid (Human+AI) (%)</b>
Equivalent Transfer	34.7%	12.4%	28.3%	38.1%
Functional Transfer	41.2%	31.8%	39.6%	42.9%
Partial Transfer	17.6%	33.4%	21.7%	14.4%
Failed / Calque	6.5%	22.4%	10.4%	4.6%

A combination of human and AI is the most efficient strategy (81.0% functional/adequate/equivalent transfer) in contrast to professional human translators (75.9%), followed by community crowd translators (67.9%) (Koehn & Knowles, 2017; Yu & Yao, 2026). NMT models show the largest percentages of partial transfers (33.4%) and unsuccessful translations/calques (22.4%) (Ghaffar, 2024; Crane et al., 2013) especially when dealing with culturally dense expressions. However, one can state that NMT systems work successfully with conceptual idioms having clear etymology (61.3% functional transfer) and poorly with historical-mythological idioms (28.7% adequate transfer) (Spasić et al., 2017).

## 6 Internet Services as Mediating Agents

### Neural Machine Translation and Idiomatic Processing

Internet-based NMT technologies namely DeepL, Google Translate, and Microsoft Translator have completely reshaped the domain of literary translation (Dobrovolskij & Piirainen, 2021). Philologically speaking, NMT technologies are unaware of the etymology of words; idiomatic meanings generated by NMT technologies are based on statistical associations in training datasets; consequently, idioms rarely or not used in literature or non-standard speech genres will be significantly underrepresented (Koehn & Knowles, 2017; Özateş et al., 2022). Culturally speaking, NMT technologies are not conscious of cultural connotations; when such a technology recognizes an idiomatic phrase, it does not have the necessary cultural knowledge to find an appropriate idiomatic expression in the target language (Spasić et al., 2017).

It was recently found out that NMT idiomatic competence is especially weak concerning culturally intensive idioms (Özateş et al., 2022). Indeed, Zhu & Xia show that language models trained with huge internet-scale datasets still do not develop sufficient culture-driven idiomatic competence; this finding is fully in line with the DPLM assumption that cultural connotational density rather than surface-level semantic complexity is the strongest predictor of automated system failure.

### Digital Literary Communities and Collaborative Translation

Web-based communities of practice ranging from translation forums catering to specific languages to multilingual literary discussion groups have become important stakeholders within the realm of informal translation practices (Yu & Yao, 2026). There were examples of online translation communities in the corpus analysis that showed a high degree of idiomatic competence in their crowd-translations, surpassing the quality of commercially produced professional translations. In at least one recorded instance, an Arabic novel with a culturally loaded Levantine dialect idiom was better translated by a fan into English via a literary subreddit than it had been professionally in a commercially released English translation, the web-based translator being a native speaker of Levantine dialects.

Thus, the important point made clear by this discovery is the ability provided by internet-based services to access scattered cultural competencies (Yu & Yao, 2026). The linguistic-cultural knowledge needed for efficient idiomatic translation constitutes a kind of specialized cultural knowledge that is not

always available to professional translators but which can be accessed through web-based communities—an innovation in literary translation infrastructure (Cemile, 2025).

### **Digital Publishing Platforms and Paratextual Mediation**

Digital publishing creates a unique model of paratextual mediation of idiomatic translation. Whereas print publications feature static paratexts such as glossaries, translator's notes, and authorial interviews, the digital format allows for dynamic paratexts that can help to explain idiomatic translations through reader annotations (Martín-Gascón, 2024). The analysis of reviewer comments collected on digital platforms has shown that culturally dense idioms often attract reviewers' attention as a source of problems with comprehension. In a review where a reader says, "The novel loses some sense in Russian idioms," which appears in 18.4 percent of all English-language reviews written by readers about Russian novels in the corpus, can observe quality control of idiomatic transfer happening in real time (Yu & Yao, 2026).

## **7 Typological Model of Idiomatic Transfer Outcomes**

On the grounds of the results obtained, the present paper offers a five-category typology of outcomes of idiomatic transfer in modern literary fiction based on the crossing of DPLM parameters and actual transfer. In addition to being a descriptive system, this typology can be regarded as predictive one useful for assessing translation quality and NLP systems (Anis et al., 2024; Dobrovol'skij, 2025).

Category I – Equivalent transfer: the target language idiomatic construct reproduces the meaning and cultural background of the source expression. Likely in case of idioms with high degree of etymological transparency, low density of connotations, and existing equivalents in the target language (Dobrovol'skij & Piirainen, 2021). For instance, English 'to burn bridges' / Arabic 'يحرق المراكب' / Russian 'сжигать за собой мосты' – all three involve the same conceptual metaphor and allow equivalent transfer.

Category II – Functional transfer: the target language construct executes the same pragmatic functions with other semantic meaning. The connotation is successfully transferred despite cultural anchoring differences. It is the most common successful strategy (42.1 % of adequate transfers) (Anis et al., 2024).

Category III – Partial Translation: The target language translation reflects some aspects of the source idiom meaning such as its explicit semantic meaning, however, it does not carry any cultural connotations, resulting in phrases that are pragmatically functional, but culturally poor (Qiufen et al., 2025).

Category IV – Substitutive Translation: Here, the translator substitutes the source idiom with another culturally equivalent phrase which fulfills similar pragmatic functions, but includes new target culture connotations. It is a method known as cultural domestication, which has been widely discussed in literature on translation (Cemile, 2025).

Category V – Failed Translation: Here, the idiom either gets translated in literal terms (a semantically unclear calque), deleted from the text, or substituted with a non-figurative phrase. This type of translation appears to be more frequent in the NMT of idioms high in historical-mythological density and entails the greatest cultural information loss (Ghaffar, 2024; Spasić et al., 2017).

## 8 Discussion: Implications and Applications

### Implications for Digital Humanities

The DPLM model provides an organized foundation for the construction of corpora containing annotated literary idioms with cultural metadata, which is a relatively neglected area within digital humanities infrastructure (Crane et al., 2013). These corpora can be used not only for qualitative research but also for training more culturally informed NLP models. Internet services acting as mediating agents may lead to the development of novel methods for the computational investigation of translation patterns from a large corpus of online translations, made possible by digital literary collections (Wierzbicka, 2012).

### Implications for Natural Language Processing

The systematic study of NMT error patterns concerning idiomatic transfer can provide valuable insights to NLP practitioners. The conclusion drawn from this study that the degree of cultural connotations is the main driver of NMT failures indicates that the enhancement of idioms' machine translation necessitates using culture-informed training datasets and assessment criteria, and not just scaling up neural models. The typology of NMT failures in five categories proposed by this paper represents a more fine-grained evaluation criterion than the adequate/ inadequate binary assessment currently applied in most MT evaluation frameworks (Dobrovolskij, 2025).

### Recommendations for Translation Practice

There are three practical implications of this research for the practice of literary translation in the era of digital technology. First, the tradition of etymological exploration as part of pre-translation preparation for idiom-heavy literary texts must be embraced as a methodological standard (Pollock et al., 2015; Dobrovolskij & Piirainen, 2021). Second, prior to translating, an analysis of cultural meaning fields dominating in the original text needs to be performed in order to pinpoint areas where cultural anchorage requires special attention (Piasecki & Kanclerz, 2022). Third, the translation strategy based on combining human effort and NMT-based automatic systems turns out to be the optimal choice due to its effectiveness at the intersection of speed and accuracy (Koehn & Knowles, 2017; Yu & Yao, 2026). Human evaluation plays an especially significant role in processing highly cultural idioms. Future work should test whether these findings generalize to other language pairs and genres.

## 9 Conclusion

The paper has made the case for how idiomatic expression transfer studies can benefit immensely from the combination of philological analysis with Linguocultural analysis methods. With the creation and employment of the DPLM model and the examination of the multilingual corpus of 1,847 idioms in English, Arabic, and Russian languages, the paper has shown that etymology and culture-specificity play a pivotal role in successful idiomatic transfer, while internet services provide both novel obstacles and opportunities in this regard. The main contributions of the paper are three-fold. From a theoretical perspective, it proposes a unified approach based on the use of philological depth analysis together with Linguocultural connotation mapping. On the empirical level, the paper provides rigorous findings regarding idiomatic transfer patterns using a large linguistically diverse corpus, focusing in particular on differences between online translation tools. And finally, from an applied point of view, it suggests a five-category model of transfer types and practical advice to translators and NLP experts. Directions for future research include extending the corpus to other language traditions especially East Asian and

sub-Saharan African literary traditions that remain under-represented in the study of comparative literary translation and the creation of computational methods for automatically analyzing DPLMs and measuring idiomatic transfer on a massive scale within digitally archived literary corpora. It seems that closer cooperation between the disciplines of linguistics, literary criticism, digital humanities, and computer science in efforts like those demonstrated above will prove both fruitful and essential in this networked world. A limitation of the study is that its corpus is restricted to ten novels and three language traditions, so broader generalization should be made cautiously.

## References

- [1] AbdelRaouf, H., Hasnaine, Q. R., Fouda, M. M., Fadlullah, Z. M., & Ibrahim, M. I. (2026). Paradigm shift toward distributed learning in iot intelligence: A comprehensive survey of opportunities and challenges. *IEEE Internet of Things Journal*, 13(10). <https://doi.org/10.1109/JIOT.2026.3657733>
- [2] Anis, M. Y., Farhah, E., Nugraha, R. S., Malik, A., & bin Syed Abdullah, S. N. (2024). Pragmatic equivalence in Syarah Al-Hikam: An attempt to explore values of humanism based on the translation of a religious text. *ALSUNİYAT: Jurnal Penelitian Bahasa, Sastra, dan Budaya Arab*, 7(2), 395-410. <https://doi.org/10.17509/alsuniyat.v7i2.74302>
- [3] Cemile, U. Z. U. N. (2025). A test of meaning, form, and culture in Kurmanji: An evaluation of large language models' performance. *Open Research Europe*, 5(313), 313. <https://doi.org/10.12688/openreseurope.20871.1>
- [4] Crane, G., Bamman, D., & Jones, A. (2013). ePhilology: When the Books Talk to Their Readers1. *A companion to digital literary studies*, 27-64. <https://doi.org/10.1002/9781405177504.ch2>
- [5] Damodaran., Lekshmi, R.S., Abdusamatov, M., Subbulakshmi, V., Azizova, S., & Jalilov, B. (2026). Smart Tutoring Systems Using Interactive Natural Language Understanding with Transformer-Based Models. *International Journal of Artificial Intelligence and Machine Learning*, 6(1s), 345-355. <https://doi.org/10.51483/IJAIML.6.1s.2026.345-355>
- [6] Dobrovol'skij, D. (2025). Phraseology and Figurative Language: Some Basic Concepts and Future Prospects. *Yearbook of Phraseology*, 16(1), 5-50. <https://doi.org/10.1515/phras-2025-0003>
- [7] Dobrovol'skij, D., & Piirainen, E. (2021). *Figurative language: Cross-cultural and cross-linguistic perspectives* (Vol. 350). Walter de Gruyter GmbH & Co KG. <https://doi.org/10.1515/zrph.2009.053>
- [8] Geetha, K., & Kumar, P. S. (2025). Bridging Artificial Intelligence, Linguistics, and Communication: Multidisciplinary Applications of Sentiment Analysis for Strategic Content Curation. *Bridge: Journal of Multidisciplinary Explorations*, 1(2), 69-76.
- [9] Ghaffar, A. (2024). Entrepreneurship AI automation. *Iconic Research and Engineering Journals*, 7(10), 531–542. <https://doi.org/10.64388/IREV7I10-1707805-259>.
- [10] Koehn, P., & Knowles, R. (2017, August). Six challenges for neural machine translation. In *Proceedings of the first workshop on neural machine translation* (pp. 28-39). <https://doi.org/10.18653/v1/w17-3204>
- [11] Kövecses, Z. (2017). Levels of metaphor. *Cognitive linguistics*, 28(2), 321-347. <https://doi.org/10.1515/cog-2016-0052>
- [12] Lakoff, G. (2016). Language and emotion. *Emotion Review*, 8(3), 269-273. <https://doi.org/10.1177/1754073915595097>
- [13] Martín-Gascón, B. (2024). Irony in American-English tweets. *Recent advances in multiword units in machine translation and translation technology*, 366, 197. <https://doi.org/10.1075/cilt.366.11mar>

- [14] Muralidharan, J. (2025). Usability-Centered Evaluation of NLP-Based Tools for Multilingual Language Therapy. *Journal of Intelligent Assistive Communication Technologies*, 1(1), 49-57.
- [15] Özateş, Ş. B., Özgür, A., Güngör, T., & Başaran, B. Ö. (2022). A hybrid deep dependency parsing approach enhanced with rules and morphology: A case study for Turkish. *IEEE Access*, 10, 93867-93886. <https://doi.org/10.1109/access.2022.3202947>
- [16] Piasecki, M., & Kanclerz, K. (2022, September). Non-contextual vs contextual word embeddings in multiword expressions detection. In *International Conference on Computational Collective Intelligence* (pp. 193-206). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16014-1\\_16](https://doi.org/10.1007/978-3-031-16014-1_16)
- [17] Piirainen, E. (2008) Chapter two widespread idioms and widespread idiomatic patterns: results for phraseology and related disciplines. *Contrastive Phraseology*, 13. <https://doi.org/10.1075/z.139.22pii>
- [18] Pollock, S., Elman, B. A., & Chang, K.-M. K. (Eds.). (2015). *World philology*. Harvard University Press. <https://doi.org/10.1086/684751>
- [19] Qiufen, W., Amini, M., & Tan, D. A. L. (2025). Strategies, errors, and challenges in translating Culture-Specific Items in Chinese-English literary works: A systematic review. *Jurnal Arbitrer*, 12(2), 259-273. <https://doi.org/10.25077/ar.12.2.259-273.2025>
- [20] Spasić, I., Williams, L., & Buerki, A. (2017). Idiom-based features in sentiment analysis: Cutting the Gordian knot. *IEEE Transactions on Affective Computing*, 11(2), 189-199. <https://doi.org/10.1109/taffc.2017.2777842>
- [21] Veerappan, S. (2025). Voices in code: Amplifying women’s experiences in tech through digital storytelling and interactive media frameworks. *Journal of Women, Innovation, and Technological Empowerment*, 1(1), 31-37.
- [22] Volkova, T. (2014). Translation model, translation analysis, translation strategy: an integrated methodology. *Procedia-Social and Behavioral Sciences*, 154, 301-304. <https://doi.org/10.1016/j.sbspro.2014.10.167>
- [23] Wierzbicka, A. (2012). Is pain a human universal? A cross-linguistic and cross-cultural perspective on pain. *Emotion Review*, 4(3), 307-317. <https://doi.org/10.1177/1754073912439761>
- [24] Yu, J. S., & Yao, Y. (2026). From Translation to Intelligent Knowledge Services. In *Intelligent Language Services: Theory and Practice with Large Language Models* (pp. 1-31). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-95-4632-9\\_1](https://doi.org/10.1007/978-981-95-4632-9_1)
- [25] Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in psychology*, 6, 149534. <https://doi.org/10.3389/fpsyg.2015.01631>

## Authors Biography



**Orif Avlayev** is an Associate Professor in the Department of Special Psychology at the Faculty of Pedagogy, Chirchik State Pedagogical University. He has extensive academic experience in the fields of special psychology, pedagogy, and inclusive education. His research interests include psychological development, special educational needs, and innovative teaching methodologies. He has contributed to various academic publications and research initiatives aimed at improving educational practices and student well-being. Dr. Avlayev is actively involved in mentoring students and promoting interdisciplinary research in education and psychology. He is based in Tashkent, Uzbekistan.



**Durdona Lutfullaeva** is affiliated with the International Islamic Academy of Uzbekistan. Her academic interests focus on education, language studies, and interdisciplinary research in the social sciences. She has been actively engaged in scholarly activities, including research, teaching, and academic collaboration. Her work contributes to the advancement of higher education and innovative learning approaches. She is committed to fostering academic excellence and supporting student development through research-oriented teaching practices. Based in Uzbekistan, she continues to participate in academic and professional initiatives at both national and international levels.



**Jasur Ziyamukhamedov** is affiliated with the Tashkent State University of Oriental Studies. His academic interests include oriental studies, language education, cultural studies, and interdisciplinary research in the humanities. He has participated in various academic and research activities aimed at promoting linguistic and cultural understanding. His work emphasizes innovative educational practices and scholarly collaboration in higher education. He is also involved in guiding students and contributing to research initiatives within his field of specialization. He is based in Tashkent, Uzbekistan.



**Dilorom Yuldasheva** is a Professor in the Department of Uzbek Linguistics and Journalism at Bukhara State University. She specializes in Uzbek linguistics, journalism, and language education, with a strong interest in linguistic research and communication studies. Her academic work focuses on the development of language teaching methodologies, discourse studies, and media communication. She has contributed to numerous scholarly publications and research projects in the fields of linguistics and journalism. Professor Yuldasheva is actively engaged in mentoring students and supporting academic research initiatives in higher education. She is based in Bukhara, Uzbekistan.



**Mashkhura Yuldasheva** is an Associate Professor in the Department of Russian Language and Literature at Bukhara State Pedagogical Institute. Her academic expertise includes Russian language teaching, literature studies, and modern pedagogical approaches in language education. She has been actively involved in research related to linguistics, literary analysis, and foreign language instruction. Her scholarly contributions support the advancement of effective teaching methodologies and cross-cultural communication studies. She is also dedicated to mentoring students and participating in academic and professional development activities. She is based in Bukhara, Uzbekistan.



**Xayriniso Suvanov** is affiliated with the Department of English Language Theory and Methodology in the Faculty of Philology at Jizzakh State Pedagogical University. Her academic interests include English language teaching, applied linguistics, and modern methodologies in foreign language education. She has participated in research and academic activities focused on improving language learning and teaching practices. Her work emphasizes communicative approaches, language skills development, and innovative pedagogical strategies. She is actively involved in supporting students' academic growth and contributing to scholarly initiatives in philology and language education. She is based in Jizzakh, Uzbekistan.



**Diyora Tashkenbayeva** is affiliated with the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Her academic interests include information technologies, digital education, and interdisciplinary research in technology-enhanced learning. She has been actively involved in academic and research activities related to innovation in higher education and modern communication technologies. Her work focuses on integrating technological solutions into teaching, learning, and professional development. She also contributes to scholarly initiatives and collaborative research projects within the field of information and communication technologies. She is based in Tashkent, Uzbekistan.