

Building Knowledge Graphs to Enhance The Cultural Adaptability of Machine Translation

Yang Jingdong¹, Mao Ting²

¹*School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;*

²*School of Foreign Languages and Cultural Communication, Shanghai Polytechnic University, China;*

Abstract: Building knowledge graphs to enhance the cultural adaptability of machine translation is a crucial approach to improving cross-cultural communication and translation quality. Traditional machine translation systems typically rely on vocabulary and grammatical rules but struggle with handling cultural differences inherent in language. By integrating cultural knowledge graphs, translation systems can identify and consider the cultural context between the source and target languages, thereby producing translations that align more closely with the cultural norms of the target language. For example, when translating references to holidays, place names, or historical figures, the system can adjust the translation to avoid misunderstandings and inaccuracies. Knowledge graphs, which consolidate vast amounts of cultural information—such as customs, traditions, festivals, and social contexts—help translation systems make more precise decisions in a multicultural environment. Moreover, knowledge graphs also support personalized translations. Depending on the user's geographical location, language preferences, and cultural background, the translation system can adjust its output to ensure that the translation meets the cultural expectations of specific audiences. For instance, the system can provide translations that reflect the cultural relevance of holidays for users in different regions, enhancing the translation's relevance and acceptability. In summary, the integration of knowledge graphs not only improves the accuracy and adaptability of machine translation but also facilitates smoother cross-cultural communication. With ongoing advancements in technology, future machine translation systems will better understand and convey cultural nuances, providing users worldwide with more natural, accurate, and meaningful translations.

Keywords: Knowledge Graphs, Machine Translation, Cultural Adaptability

1. Introduction

Translation models leverage machine learning (ML) and natural language processing (NLP) techniques to automatically convert text between languages. The evolution of these models has been driven by the advancement of deep learning algorithms, particularly neural networks such as the Transformer model, which has significantly enhanced translation quality and fluency. Rule-Based Machine Translation (RBMT) depends on predefined linguistic rules and dictionaries to map the syntax of a source language to a target language. Although effective for certain language pairs with well-established grammar, RBMT requires extensive manual rule crafting and struggles to handle the complexity and diversity of languages. It is also less adaptive when facing variations in linguistic nuances and idiomatic expressions (Hutchins & Somers, 1992). Statistical Machine Translation (SMT) relies on statistical models and large bilingual corpora to generate translations. This technique estimates the most likely translation by analyzing the frequency of word pairs and sentence structures found in parallel corpora. SMT models, while effective, can sometimes produce awkward or unnatural translations due to their reliance on probabilities and lack of deep linguistic understanding (Koehn et al., 2003). Neural Machine Translation (NMT) employs deep learning techniques, particularly neural networks, to model the translation process as an end-to-end system. The Transformer model, which underpins most

modern NMT systems, excels in capturing long-range dependencies and complex relationships between source and target languages. This results in translations that are more accurate and fluent compared to earlier models. NMT represents the current state-of-the-art in machine translation (Vaswani et al., 2017). While significant advances in deep learning and large-scale data processing have greatly enhanced the accuracy and fluency of translation models, challenges remain, especially for languages with complex syntactical rules or idiomatic expressions.

2. Challenges in Translation Models

One of the key challenges in machine translation is resolving the ambiguity of words that can have different meanings depending on the context. Properly disambiguating these words is crucial for producing accurate translations. Current models struggle to fully capture and correctly interpret contextual nuances (Mikolov et al., 2013). Also, languages often differ in their grammatical structure, word order, and syntax. For example, the subject-verb-object order of English sentences may be quite different from the subject-object-verb order in languages like Japanese. Handling these structural differences without losing meaning or fluency in the translation is a continuing challenge (Bahdanau et al., 2014). In addition, idiomatic expressions, cultural references, and localized phrases may lack direct equivalents in other languages, complicating translation. NMT models are increasingly capable of handling these issues, but they are still limited in some cases by the availability and diversity of training data.

3. Cross-Cultural Adaptability of Machine Translation

The cross-cultural adaptability of machine translation (MT) refers to the ability of translation systems to effectively operate across different cultural and linguistic contexts. This adaptability is increasingly important as global communication intensifies, requiring translation systems to handle not only linguistic variations but also cultural nuances embedded in language. The literature on this topic spans various areas, including the impact of cultural differences on translation quality, the challenges of cross-cultural communication, and the strategies that machine translation systems use to adjust to these differences.

One major aspect of cross-cultural adaptability is the influence of cultural differences on translation accuracy and fluency. Nivre (2006) emphasizes how culture-specific elements such as idioms, slang, and cultural references can affect the quality of translation. Similarly, Hutchins (2004) explores the limitations of MT systems when translating between languages that have distinct cultural contexts. These challenges arise because language is inherently intertwined with culture, and machine translation systems must navigate these complexities to produce accurate and contextually appropriate translations.

Römer (2004) and Gallois and Callan (1997) further investigate the role of culture in shaping language use and communication practices. Römer argues that cultural concepts and habitual expressions play a significant role in how languages evolve and how meaning is conveyed. Cultural background influences communication, which is particularly critical when translating between languages with starkly different cultural frameworks. These studies underscore the need for machine translation systems to not only understand linguistic structures but also the cultural contexts in which these structures operate.

To address these challenges, various strategies for cultural adaptation in machine translation have

been proposed. Koehn (2009) discusses statistical machine translation (SMT) and how it uses large datasets to adapt to language and cultural differences. Dwivedi et al (2024) introduce specific strategies aimed at enhancing MT's cross-cultural adaptability, such as localizing content to better suit target cultures. The advent of neural machine translation (NMT) has further advanced these efforts. Conia et al (2016) introduces XC-Translate benchmark and KG-MT method, significantly improving entity name translation by integrating multilingual knowledge graphs into neural models. The Transformer model (Vaswani et al., 2017) introduced new techniques for processing cultural data, significantly enhancing the flexibility and accuracy of NMT in diverse cultural settings.

Moreover, localization in translation, as discussed by Esselink (2000) and Baker (2018), is crucial for adapting translations to local cultural norms. Software localization, for instance, integrates MT with cultural adaptation to ensure that translated content is contextually appropriate for different audiences.

Finally, the evaluation of cultural adaptability in MT systems remains a key area for future research. Naveen an Trojovský (2024) highlights challenges in achieving accurate machine translation, focusing on context, ambiguity, idioms, cultural nuances, and domain-specific terms, to improve global communication.. Bapna and Firat (2019) focus on how neural networks can be fine-tuned to adapt to specific linguistic elements, presenting strategies to enhance MT performance in cross-cultural contexts.

In conclusion, the cross-cultural adaptability of machine translation is an interdisciplinary field that involves linguistics, cultural studies, and computer science. Despite significant progress, particularly with the advent of NMT and deep learning, challenges remain in translating culture-specific content. Future research will likely focus on refining MT systems to better handle cultural references, idiomatic expressions, and localization, ensuring that translations are not only linguistically accurate but also culturally appropriate.

4. The Role of Cultural Knowledge Graphs in Enhancing the Cultural Adaptability of Machine Translation (MT)

The integration of cultural knowledge graphs into machine translation (MT) is a critical development that enables MT systems to navigate the intricacies of language and culture. These knowledge graphs play a vital role in ensuring that translations are not only linguistically accurate but also culturally appropriate. The following aspects illustrate how cultural knowledge graphs enhance the cultural adaptability of MT:

4.1. Understanding Cultural Contexts

Machine translation is not simply about translating words; it is about conveying meaning, nuances, idiomatic expressions, and culturally specific references. Traditional MT systems often struggle to handle culturally embedded concepts, resulting in mistranslations. Cultural knowledge graphs structure information regarding customs, traditions, social norms, and historical backgrounds across languages and cultures. By organizing this information, knowledge graphs allow MT systems to recognize and preserve cultural contexts, thereby improving the system's understanding of the source language and producing more culturally sensitive translations. This can prevent issues such as the inappropriate use of idiomatic expressions or culturally sensitive terms.

4.2. Eliminating Linguistic Ambiguities

Many words and phrases have multiple meanings, with these meanings heavily dependent on cultural context. In MT, words that appear straightforward in one language can carry additional, culture-bound meanings in another. For instance, a phrase such as "break a leg" may be understood as an idiomatic expression in English, but in other cultures, it might be translated literally, leading to confusion. Cultural knowledge graphs can provide mappings between languages, helping MT systems identify and resolve such ambiguities by offering contextual information about how words and phrases are used in different cultures (Chen et al., 2020). This is essential for ensuring that translations are both accurate and culturally adaptable.

4.3. Cross-Cultural Language Conversion

Expressions, proverbs, or even references to public holidays can be deeply tied to specific cultural contexts, making direct translation impossible. For example, while "Christmas" is universally recognized in many cultures, its significance and specific traditions may vary across regions. Cultural knowledge graphs aid MT systems by organizing and mapping out cross-cultural knowledge, such as regional festivals, customs, and holidays. This enables the system to adjust references according to the cultural context of the target language. For instance, the term "Chinese New Year" might be better understood and translated into a Western language as "Spring Festival" when referring to the Chinese tradition, enhancing both clarity and cultural relevance.

4.4. Cultural Sensitivity and Contextual Adjustment

Cultural sensitivity is an essential component of effective machine translation. What is considered acceptable or polite in one culture may be seen as offensive in another. Cultural knowledge graphs assist MT systems in identifying sensitive or taboo topics and adjusting translations accordingly. This includes not only word choices but also sentence structure, tone, and even the level of detail. For example, content related to religion or politics may need to be carefully rephrased to avoid cultural insensitivity. Such contextual adjustments ensure that translations are not only linguistically accurate but also appropriate for the target audience.

4.5. Enhancing Contextual Awareness in Machine Translation

Cultural knowledge graphs contribute to the contextual understanding of MT systems. The meaning of words or phrases can shift depending on the cultural background, and cultural knowledge graphs help systems differentiate these variations. For example, the word "respect" may carry different connotations depending on whether the culture is collectivist or individualist. By embedding such cultural nuances into the translation process, MT systems can generate translations that are more natural and fluent in a given cultural setting.

4.6. Improving Accuracy in Specialized Fields

Specialized fields like law, medicine, and business have their own terminology and cultural expectations, which can impact translation accuracy. For instance, legal systems and medical practices vary significantly across cultures, and a translation that works in one context may be

completely inappropriate in another. By embedding domain-specific cultural information into knowledge graphs, MT systems can navigate these fields with greater precision. In legal translations, for example, understanding the differences between common law and civil law systems can prevent misinterpretations. In medical translations, cultural knowledge graphs help address region-specific medical practices and terminologies, enhancing both the accuracy and relevance of translations in these specialized domains.

Cultural knowledge graphs provide a critical framework for improving the cultural adaptability of machine translation. By embedding rich cultural information—ranging from idiomatic expressions to regional customs—into MT systems, these graphs help ensure that translations are contextually appropriate, sensitive to cultural nuances, and aligned with the expectations of the target audience. This not only enhances translation quality but also fosters smoother and more effective cross-cultural communication.

5. Construction of a Cultural Knowledge Graph

To enhance the cultural adaptability of machine translation, the construction of a cultural knowledge graph is a multi-dimensional, systematic endeavor. This process involves several critical stages, including the development of the knowledge graph structure, data collection, model optimization, and more. The following provides a comprehensive analysis and implementation framework for this task:

5.1. Defining the Structure of the Cultural Knowledge Graph

A cultural knowledge graph systematically organizes culture-related information, encompassing three main components:

5.1.1 Cultural Entities

Cultural entities encompass a wide range of dimensions, each representing unique aspects of human society and experience. Geography, for instance, includes countries, cities, and landmarks, which shape the identity of a place and its people. History contributes through significant events and influential figures that have impacted the course of civilization. Traditions play a vital role in reflecting the customs and festivals that define cultural practices. Art forms such as music, dance, and painting further embody the aesthetic and emotional expressions of a community. Customs and etiquette guide the social behavior within regions, offering insight into local values and practices. The people associated with these entities, from historical figures to modern celebrities, help bring cultural narratives to life. Each cultural entity is characterized by specific attributes, such as the dates and manners of holiday celebrations, which reflect the historical context and significance of those customs. Detailed event information, including the time, location, and key figures involved, provides a deeper understanding of how cultural practices evolve and maintain their relevance over time. Together, these dimensions shape the richness and diversity of human culture, offering both a window into the past and a bridge to the present.

5.1.2 Cultural Relationships

These relationships illustrate the intricate interactions and interconnections among cultural entities within historical, social, and cultural contexts. For instance, the "French Revolution" and "Napoleon" are deeply intertwined; the revolution set the stage for Napoleon's rise to power,

fundamentally altering France's political and social structure. Similarly, festivals and customs reflect the values and traditions of a culture, such as the practice of "gift-giving" during "Christmas," which symbolizes love, generosity, and family bonds, marking a significant part of many cultures' holiday traditions. Art forms also demonstrate a close connection with regional culture; for example, "Peking Opera" is not just a performance art but an essential component of Beijing's cultural identity, reflecting the city's history, language, and aesthetic sensibilities. These relationships show how historical events, social practices, and artistic expressions are often linked, with each influencing and shaping the other over time. Understanding these interconnections allows us to appreciate the depth and complexity of cultural evolution, revealing how deeply history, customs, and art are woven into the fabric of society.

5.1.3 Cultural Attributes

Cultural attributes offer valuable insights into the unique characteristics of specific cultural entities, enriching our understanding of their significance and evolution. For example, the significance of a holiday extends beyond its celebratory aspects, often reflecting deep cultural values, historical milestones, or religious beliefs. An example from China is Chinese New Year (also known as the Spring Festival). While the holiday is widely celebrated with fireworks, family reunions, and feasts, its significance goes far beyond these celebratory aspects. Chinese New Year marks the beginning of the lunar calendar and is deeply rooted in cultural values such as family unity, respect for ancestors, and the renewal of life. The holiday is associated with various traditions, including the cleaning of homes to sweep away bad luck and the giving of red envelopes for good fortune. It also reflects historical milestones, such as the myth of the Nian monster and the tradition of driving away evil spirits. Overall, Chinese New Year is a time for honoring cultural heritage, fostering family bonds, and hoping for a prosperous year ahead. Similarly, the origin and evolution of particular customs also provide important cultural context. The tradition of "tea-drinking" in British culture, for instance, evolved from the 17th century as a social ritual, growing into a symbol of hospitality and British refinement. These cultural attributes help us appreciate the deeper meanings behind customs, events, and practices, shedding light on how they shape identities and continue to influence societies across generations.

5.2. Collecting and Integrating Cultural Data

The development of a cultural knowledge graph relies on gathering a wide range of cultural data from multiple sources to ensure accuracy and depth. Literature forms a core component, providing historical texts, cultural research, and academic papers that offer a scholarly perspective on cultural practices, beliefs, and histories. These sources help contextualize cultural phenomena over time. Online resources like Wikidata, DBpedia, and Wikipedia serve as invaluable platforms for aggregating structured and unstructured cultural knowledge. These open knowledge repositories offer extensive, crowd-sourced information, including details about cultural events, figures, and artifacts, which can be linked and categorized within the graph. Additionally, expert knowledge is crucial for refining and validating the data. Insights from cultural experts, such as anthropologists, historians, or local practitioners, provide nuanced details on topics like local festivals, rituals, and traditions. These expert annotations help fill in gaps that raw data or general resources might miss. By integrating these diverse data sources, a cultural knowledge graph can provide a comprehensive, dynamic view of cultural entities and their interconnections across time and space.

5.3. Constructing a Cultural Corpus

To improve machine translation's ability to handle cultural nuances, developing a cross-linguistic and cross-cultural corpus is crucial. This corpus should encompass culturally-specific vocabulary, which includes terms deeply rooted in particular cultural contexts, such as "Spring Festival" in Chinese or "Thanksgiving" in English. These terms often carry meanings that extend beyond direct translation, requiring contextual sensitivity. Contextual annotations should also be integrated, linking cultural references to their linguistic context, ensuring that background information is considered during translation. For instance, understanding the significance of a festival or historical event can determine how a term is best translated into another language. Lastly, cross-cultural examples play a vital role by including texts that showcase the impact of cultural differences on translation. Analyzing such examples helps identify translation challenges, revealing how culture shapes meaning and suggesting strategies for better handling of these challenges. By combining these elements, the corpus would provide valuable resources to enhance machine translation's cultural awareness, improving both accuracy and appropriateness in cross-cultural communication.

5.4. Establishing Multilingual Mapping

Given that cultural concepts often lack direct equivalents across languages, establishing multilingual mappings is essential for accurate and contextually appropriate translations. Cross-language comparisons are key in this process. For instance, the term "春节" in Chinese is translated as "Chinese New Year" in English. However, the cultural significance behind the term—such as the holiday's historical roots and the tradition of family reunion—must be conveyed in the translation. This ensures that the target audience understands not just the word, but its deeper meaning and context. Similarly, vocabulary and idiomatic expression mapping is crucial for terms that carry cultural weight. A term like "饺子" (Chinese dumplings) may be translated as "dumplings" in English, but this misses its cultural importance, particularly during Chinese New Year when dumplings symbolize family unity and prosperity. A translation system should account for such cultural nuances to provide accurate, culturally sensitive translations that resonate with the target audience, reflecting not just linguistic, but also cultural equivalences.

5.5. Building a Culturally Adaptive Translation Model

The cultural knowledge graph can be leveraged to enhance machine translation's cultural adaptability in the following ways:

5.5.1 Context-Aware Translation

The system queries the knowledge graph to obtain relevant information based on the cultural context of the input text. For instance, when translating "春节," the system may opt for "Chinese New Year" rather than "Spring Festival" to avoid potential cultural misunderstandings.

5.5.2 Entity Recognition and Cultural Information Insertion

The translation system identifies cultural entities in the source language (such as holidays, historical figures, or place names) and adjusts the translation based on the cultural norms of the target language. For example, "Santa Claus" may be translated as "Father Christmas" or "Santa Claus," depending on the cultural context.

5.5.3 Diverse Translation Outputs

For sentences with cultural differences, the system can generate multiple translation options, selecting the most appropriate one based on context and the cultural habits of the target language. Certain proper nouns, for instance, may have different translation methods depending on the context.

5.6. Continuous Update and Optimization of the Knowledge Graph

To ensure the cultural knowledge graph remains relevant and accurate, it must be continuously updated through various methods. First, a user feedback mechanism is essential, allowing users to provide insights from real-world translation applications. This feedback helps identify gaps or inaccuracies, ensuring the knowledge graph reflects evolving cultural contexts. Second, automated knowledge graph updates can be facilitated using web crawlers and natural language processing technologies. These tools enable the automatic collection of new cultural data from diverse sources, including news outlets, social media, and academic papers, which are then integrated into the graph. Lastly, cross-disciplinary collaboration is crucial for maintaining the graph's relevance. Regular reviews by cultural scholars, translation experts, and specialists ensure the data is both accurate and timely, accounting for shifts in language, culture, and societal trends. By combining these approaches, the knowledge graph can remain dynamic and responsive to the ever-changing nature of language and culture.

5.7. Practical Applications

Integrating the cultural knowledge graph with machine translation systems brings significant benefits to real-world applications. Cross-language cultural communication is enhanced by culturally adaptive translations in fields like news reports and literary works, enabling target-language readers to grasp the cultural nuances of the original text. Personalized translation tailors content to a user's geographic and linguistic background, adjusting translations to fit cultural norms. For instance, festival-related terms can be customized depending on the user's country of origin, ensuring relevance. Lastly, in multi-cultural communication platforms, such as international conferences, culturally adaptive translations help minimize misunderstandings, fostering more effective dialogue and collaboration across diverse cultural contexts. This integration enhances accuracy, cultural sensitivity, and overall communication efficiency.

6. Conclusion

The construction of a cultural knowledge graph significantly enhances the ability of machine translation systems to recognize and adapt to cultural nuances within the source text. This adaptation improves both translation accuracy and fluency while fostering greater cultural understanding. Building such a graph requires a comprehensive effort, spanning data collection, corpus construction, cross-lingual mapping, translation model optimization, and ongoing updates and revisions. Ultimately, this approach enables high-quality, culturally sensitive translations in diverse, multi-cultural contexts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Acknowledgments

This research is funded by the Undergraduate Teaching Research and Reform Project of **University of Shanghai for Science and Technology** (JGXM24124)

References

- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473.
- Baker, M. (2018). *Translation and Conflict: A Narrative Account*. Routledge.
- Gallois, C., & Callan, V. J. (1997). *Communication and culture: A guide for practice*. John Wiley & Sons Inc.
- Bapna A. and Firat O. 2019. [Simple, Scalable Adaptation for Neural Machine Translation](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1538–1548, Hong Kong, China. Association for Computational Linguistics.
- Esselink, B. (2000). *A Practical Guide to Localization*. John Benjamins Publishing.
- Hutchins, W.J. (2006). *Machine Translation: History*.
- Hutchins, W. J. & Somers, H. L. (1992). *An Introduction to Machine Translation*. Academic Press.
- Koehn, P. (2009). *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Koehn, P., Franz J. Och, and Daniel Marcu. (2003). In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 127–133.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Neural Information Processing Systems (NIPS) 2013*.
- Römer, U. (2024), Usage-Based Approaches to Second Language Acquisition Vis-à-Vis Data-Driven Learning. *TESOL J*, 58: 1235-1245. <https://doi.org/10.1002/tesq.3278>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017)*, 5998-6008.
- Conia S., Lee D., Li M., Minhas F. U., Potdar S., and Li Y.. 2024. [Towards Cross-Cultural Machine Translation with Retrieval-Augmented Generation from Multilingual Knowledge Graphs](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 16343–16360, Miami, Florida, USA. Association for Computational Linguistics.
- Dwivedi, S., Ghosh, S. & Dwivedi, S. (2024). Navigating Linguistic Diversity: In-Context Learning and Prompt Engineering for Subjectivity Analysis in Low-Resource Languages. *SN COMPUT. SCI*, 5, 418. <https://doi.org/10.1007/s42979-024-02770-z>
- Naveen P., Trojovský P. (2024) Overview and challenges of machine translation for contextually appropriate translations, *iScience*, Volume 27, Issue 10.