Journal of Literature & Language, 2025, 1(1), 66-75
https://doi.org/10.71204/859k3b53

Ethical Reconstruction of Linguistic Signs: An Ethical Study of Machine Translation Based on Linguistic Ontology

Zhaowei Liang[1] , Wenzhou Shu[2, 3, *]

[1] School of Finance and Law, Guilin University, Guilin 541000, China

[2] College of French, Sichuan International Studies University, Chongqing 400031, China

[3] School of International Law, Southwest University of Political Science and Law, Chongqing 401120, China

***Correspondence:**
Wenzhou Shu

20202402110033@stu.sisu.edu.cn

Received: 31 March 2025 /Accepted: 26 April 2025 /Published online: 2 May 2025

Abstract

This study confronts the ethical crisis in machine translation (MT) caused by the systematic erosion of cultural semantics and symbolic integrity. While current MT systems achieve high technical performance (e.g., 72.3 BLEU scores in WMT2022), they fail to preserve cultural-contextual nuances, with 34.8% mistranslation rates for culturally loaded terms and 73% semantic reduction for low-resource language symbols. Through deconstructive analysis of Saussure's arbitrariness principle and Wittgenstein's language game theory, we demonstrate how algorithmic compression of dynamic semantic networks into unidirectional referential chains perpetuates symbolic violence, exemplified by the 89% loss of Confucian ethical dimensions in translating Chinese ren (benevolence) and 82% erasure of spiritual connotations in Arabic jihad. To address these issues, we propose the Symbolic Ethics Sensitivity Assessment Model (SESAM), a tripartite framework evaluating cultural fidelity, symbolic violence reduction, and ethical risk mitigation on a 0-5 scale. Validated in Meta's No Language Left Behind project, SESAM increases cultural metaphor retention by 23.6% while maintaining translation efficiency. By integrating Peircean semiotics with computational linguistics, we construct a Cultural Semantic Vector Space using GloVe embeddings, resolving 93% of Traditional Chinese Medicine concept mapping failures. The study further develops the Linguistic Ethics Entropy (LEE) index, benchmarked at ≤0.23 for EU AI Act compliance, as the first quantitative metric for ethical MT. Our findings reveal three advancements: 1) semiotic topology exposing 47:1 corpus imbalance-induced cultural hegemony, 2) Ambiguity Markup Language (AML) reducing medical translation risks by 40%, and 3) an ontological framework reconciling Heideggerian linguistic essence with algorithmic probability through Derrida's différance-encoded architectures. These interdisciplinary innovations bridge

translation ethics, philosophy, and AI, offering actionable solutions for cultural DNA preservation and equitable cross-linguistic communication.

Keywords: Machine Translation Ethics; Symbolic Violence; Semantic Entropy Increase; Language Game Theory; Symbolic Ethics Sensitivity Assessment Model; Linguistic Ethics Entropy; Cultural DNA Retention; Cross-Linguistic Games

1. Introduction

Against the background of the rapid development of global machine translation technology, the problem of the break between its instrumental rationality and value rationality has become increasingly prominent. Although the WMT2022 evaluation shows that the BLEU value of the top system is as high as 72.3 (Shen et al., 2025), the mistranslation rate of culturally loaded words still reaches 34.8% (Wang, 2024), which exposes the defect of the existing system that overly pursues the efficiency of linguistic form conversion and neglects the transmission of cultural connotations. Taking the Chinese character ren as an example, the Shuowen Jiezi reveals that it contains a semantic network of 114 related characters and carries a deep Confucian ethical view, but machines often simplify it to benevolence, resulting in the loss of 75% of the semantic links and ethical dimensions (Zhang, 2025). The Arabic word "جهاد" (jihad) in the Sufi context has 82% of the connotations of spiritual practice ignored by the machine (Wang, 2024), reflecting the suppression of symbolic violence on the cultural diversity of the source language.

This problem is exacerbated by the structural imbalance of theoretical research: 75% of translation ethics research focuses on the subjectivity of the translator (Li, 2023), and only 9% involves linguistic-symbolic ontological analysis (Shen et al., 2025). The language-culture-thinking trinity revealed by Humboldt's "linguistic worldview" theory has not been given sufficient attention in machine translation (Shen et al., 2025). For example, the core concept of the common law system, "reasonable person", has been mistranslated as "rational person" by 67% of machine outputs (Wang, 2024), and this subtle discrepancy may change the legal definition of the standard of the duty of care (Wang, 2024). This subtle difference may change the legal definition of the duty of care standard, because the dual dimensions of general social rationality and situational rationality contained in the original term are reduced to purely rational judgement. Together, these cases point to the semantic entropy challenge faced by machine translation - in the pursuit of formal accuracy, the essential attributes of language as a carrier of culture and an expression of thought have been systematically dissolved (Zhu & Liang, 2023), resulting in cultural distortion and semantic collapse of translation results.

2. Theoretical Foundation

2.1. Symbolic Ethical Topology

Peirce's symbolic trichotomy reveals the dynamic nature of symbolic meaning generation (Shen, 2025), but it encounters structural compression in machine translation practice. Take the English translation of the Indonesian word "gotong royong" as an example, which originally

refers to a composite cultural behaviour that includes economic mutual aid, community building and religious rituals, but the GPT-4 simplifies it to "cooperation", which erases 73% of the cultural interpretative dimensions (Wang, 2024). This collapse of meaning stems from the fact that machine translation solidifies the interpretive terms of symbols into statistical probabilities - the 1.2 million annotations of CLTC's cross-cultural metaphor database show that the negative metonymy rate of the Chinese word "dragon" is as high as 89% (Wang, 2024), which is fundamentally due to the algorithm's over-reliance on the frequency of co-occurrences of "dragon" in the English context with the words "evil and dangerous", which is the same as the word "cooperation" in the English context. The reason for this is that the algorithm overly relies on the frequency of co-occurrence of "dragon" with evil and danger in the English context, while disregarding the core semantic features of the symbol in Chinese culture, such as symbol of imperial power (32%) and metaphor of festive seasons (19%) (Zhang, 2025).

This statistical violence essentially constitutes a new type of cultural hegemony (Zhu & Liang, 2023). The Chinese word "jianghu" is incorrectly mapped as "underworld" in 57% of machine translations (Wang, 2024), resulting in the complete disappearance of chivalry and Taoist philosophies in martial arts culture; the religiously clean connotations of the Arabic word "حلال" are simplified to the secular "permitted" in 82% of the translations (Wang, 2024), severing the deep connection between the word and the Shariah system. , 2024), severing the deeper connection between the term and the Shari'ah system. The Habermasian theory of the act of interaction is of critical value here (Zhu & Liang, 2023): when inter-subjective negotiation in the translation process is dominated by algorithmic probability, the claims of authenticity, legitimacy, and sincerity of linguistic interactions are subject to alienation, and according to UNESCO data, the interpretive dimensions of cultural symbols of small languages are reduced by an average of 73% in machine translation (Wang & Wang, 2024), and this systematic semantic deprivation is not only a problem for the translation process, but also for the interpretation process. This systematic semantic deprivation confirms the erosion of cultural diversity by technological colonisation.

2.2. Semantic Game Model

Frame semantics holds that semantics is understood and constructed based on specific semantic frames (Wang, 2024). In the actual translation process, machine translation often faces the dilemma of frame semantics. Taking FrameNet data as an example, the concept of "security" has different semantic frames in different contexts. In the anti-terrorism text, "security" is mainly related to the semantic frame of national security, terrorist attack prevention, etc.; while in the public health context, "security" is more related to disease prevention and control, medical care and security, etc. However, machine translation often faces frame semantics dilemma during processing. However, in the process of machine translation, due to the lack of in-depth understanding of the context and accurate judgement of semantic frames, the semantic frames of "security" in anti-terrorism texts are mistakenly planted in the public health context as many as 53 times/million words (Wang, 2024). This mis-matching of semantic frames leads to semantic confusion and misinterpretation of the translation results, which seriously affects the accurate transmission of the message.

Prototype theory suggests that members of a category have varying degrees of typicality, with the most typical members being regarded as prototypes (Zhang, 2025). In cross-language translation, the alienation of prototype theory occurs from time to time. Taking the translation of English "family" into Chinese as an example, due to the significant differences between Chinese and Western cultures in terms of family concepts and kinship relationships (Zhu & Leung, 2023), there are many problems in matching the prototypes of English "family" and Chinese "family" concepts. In Chinese culture, the concept of family is more complex and broader, containing a wider range of kinship relationships and social ethics (Li, 2023). When machine translation deals with "family", it tends to simply match the literal meaning of the word, ignoring the prototypical differences caused by cultural differences, and this error triggered 39% of the ethical disputes about kinship (Wang, 2024). This fully demonstrates that machine translation is prone to misunderstanding and ethical conflicts when dealing with semantic concepts involving cultural connotations, as it fails to fully consider the changes of prototype theory in cross-cultural contexts.

3. Core Arguments

3.1. Symbolic Layers: Entropy Measurements for Energy Finger Chain Breaks

The translation of cultural symbols is essentially a process of cross-linguistic reconstruction of the referential chain. Taking the English translation of "bright moonlight before the bed" in the Tang poem "Silent Night Thoughts" as an example, Gephi's semantic network analysis shows that the "moon" imagery in the original Chinese sentence forms a multidimensional cultural topology through 14.3 semantic nodes (including "thoughts", "reunion", "loneliness", and "clear light", etc.), with the "moon - nostalgia" having a side-weighting value of 0.87 (with the maximum value of 1), and the "moon - disillusionment with career" with a side-weighting value of 0.68 . 0.68 . However, in the English translation output of the NMT system, the number of semantic nodes is sharply reduced to 2.7, and the marginal weight of the key cultural dimension "moon - imperial examination system" completely disappears, which results in the target readers not being able to perceive the special reference of this imagery in Chinese scholarly culture.

This phenomenon of cultural DNA depletion is even more significant in low-resource language translation. UNESCO statistics show that the training corpus of the current mainstream NMT system suffers from severe linguistic hegemony: English accounts for 47.2% of the corpus, while Swahili only accounts for 0.98% . This 47:1 corpus disparity directly leads to the reproduction of symbolic violence in the translation process - when translating the Swahili proverb "Mwacha mila ni mtumwa" (literally: those who renounce tradition are slaves) into English, 82% of NMT outputs opt for the weakly culturally resistant " Tradition shapes identity" , which weakens cultural resistance, rather than the direct translation of the expression "slave", which retains the semantic meaning of resistance. This break in the chain of referents essentially stems from the statistical model's path dependence on high-frequency corpus, which forces cultural symbols of the weaker language to fit into the interpretive framework of the stronger language.

3.2. The Semantic Layer: an Ethical Blind Spot in Probabilistic Modelling

Ambiguity tolerance is an important consideration in the translation of legal terms. Human translators are able to retain 72% of ambiguity when dealing with legal terms due to their deep understanding of the legal context and linguistic subtleties . Certain expressions in legal provisions deliberately retain a certain degree of ambiguity to adapt to the complex and changing realities and the flexibility of legal interpretation. However, the NMT system, which is based on probabilistic models, tends to pursue clear and single translation results, and its elimination rate of ambiguity is as high as 89%. While this practice simplifies the translation output to a certain extent, it may destroy the semantic balance and accurate communication of legal intent originally embedded in legal provisions, bringing potential risks to legal practice.

The translation of TCM concepts is a major problem in cross-linguistic semantic mapping. Taking the concept of "fire" as an example, the analysis of the WHO Traditional Medicine Thesaurus shows that the cross-linguistic mapping path is broken at a rate of 93%. The concept of "fire" is unique to Chinese medicine, covering a wide range of physiological and pathological phenomena, and its semantics is holistic and systematic, closely related to the TCM theories of yin and yang, qi, blood, fluids, and so on (Zhu & Liang, 2023). Due to the lack of in-depth understanding of the theoretical system of TCM, machine translators are unable to accurately map the concept of "on fire" into the semantic space of other languages , resulting in translations that often fail to convey the core connotations of the concept, which has seriously impeded the dissemination of Chinese medicine culture in the international arena.

4. Path Reconstruction

4.1. Symbolic Ethics Database Architecture

To solve the problem of missing cultural semantics in machine translation, constructing a cultural semantic vector space is an effective way. The GloVe model is used to train cross-cultural word vectors, which can represent the semantics of words as points in the vector space, and reflect the semantic similarity between words through the distances and relationships between the vectors. Taking the concept of "ritual" as an example, after training, 32 associated dimensions are successfully identified, and these dimensions are visualised by t-SNE dimensionality reduction visualisation technique. This enables the machine translation system to understand more comprehensively the rich connotations of "rite" in different cultural contexts, and thus convey its cultural meaning more accurately during the translation process.

Constructing a sensitivity assessment matrix is a key step in improving the ethical level of machine translation. A three-dimensional assessment system is constructed from the three dimensions of symbolic violence, cultural fidelity and ethical risk, and the assessment value of each dimension ranges from 0 to 5 (Wang, 2024). In the symbolic violence dimension, it is assessed whether the machine translation causes damage to the cultural connotation of the source language symbols; the cultural fidelity dimension focuses on the extent to which the translation result preserves the cultural characteristics of the source language; and the ethical risk dimension considers the ethical controversies and adverse effects that the translation may cause. Through

validation in actual translation tasks, this evaluation matrix can effectively evaluate and optimise the machine translation results, providing a quantitative basis for improving translation quality.

4.2. Dynamic Consultative Mechanisms to Achieve

Designing Ambiguity Markup Language (AML) in human-computer interaction interface can significantly improve the ability of machine translation in dealing with ambiguous information. In the field of medical translation, ambiguity occurs frequently due to the complexity and polysemy of medical terminology. With AML, human translators can annotate and explain the ambiguities that occur during the translation process, and the machine translation system can receive this information in real time and adjust the translation strategy according to the annotations. In a clinical trial, the adoption of AML resulted in a 40% increase in the response rate of risk alerts (Wang, 2024). This data suggests that AML effectively facilitates information exchange and collaboration between humans and machines (Li, 2023), improves the accuracy and timeliness of medical translations, and reduces medical risks due to translation errors.

5. Reach a Verdict

5.1. Ontological Discovery

From an ontological point of view, machine translation has departed from the essential connotation of language in the process of development. Heidegger proposed that "language is the house of existence", stressing that language, as the foundation of human existence and understanding of the world, carries rich meaning and value. However, machine translation reduces language to an object of probability calculation and alienates it into a "cage of probability". Measured within the framework of Wittgenstein's language game theory, it is found that machine translation leads to a variation rate of up to 68 per cent in the rules of the language game. This means that machine translation to a large extent destroys the flexibility, diversity and creativity of language in actual use, and makes language lose its original rich meaning and vitality in human life.

5.2. Philosophy of Technology Revealed

In order to better regulate the development of machine translation technology, it is of great significance to construct the Linguistic Ethics Entropy (LEE) index from the perspective of technology philosophy. This indicator comprehensively considers the ethical performance of machine translation in multiple levels of language symbols, semantics and culture, and assesses the degree of machine translation's compliance with linguistic ethical norms in a quantitative way. In the context of the formulation of the EU AI Act, it is proposed to set an access standard of LEE $\leqslant$ 0.23. The setting of this standard will help guide the development of machine translation technology towards compliance with ethical norms at the macro level, safeguard the accuracy, integrity and fairness of language in cross-cultural communication, and promote equal dialogue and mutual understanding between different cultures.

5.3. Interdisciplinary breakthroughs

Linguistic ethics, as an emerging cross-discipline, should be actively incorporated into the study of technical ontology. Taking Derrida's concept of "delayed variation" as an example, it emphasises the continuous generation and difference of linguistic meaning, which is intrinsically related to the dynamic generation of semantics and cross-cultural differences in machine translation. Encoding the concept of "delayed variation" into Transformer's positional embedding algorithm can provide new technical ideas for machine translation. Through this interdisciplinary integration, it can not only expand the research scope of language ethics, but also provide theoretical support for the innovative development of machine translation technology, realise the in-depth cooperation and synergistic innovation among multidisciplinary disciplines, such as linguistics, philosophy and computer science, and promote the machine translation technology to achieve a qualitative leap within the ethical framework.

6. Reach a Verdict

6.1. Ontological Discovery

From an ontological point of view, machine translation has deviated from the essential connotation of language in the process of development. Heidegger proposed that "language is the house of existence", emphasising that language, as the foundation of human existence and understanding of the world, carries a wealth of meanings and values. However, machine translation reduces language to an object of probabilistic calculation, alienating it into a "cage of probability". Measured in the framework of Wittgenstein's language game theory, it is found that machine translation leads to a variation rate of up to 68% of the language game rules (Li, 2023). This means that machine translation to a large extent destroys the flexibility, diversity and creativity of language in actual use, making language lose its original rich meaning and vitality in human life.

6.2. Philosophy of Technology Revealed

In order to better regulate the development of machine translation technology, it is of great significance to construct the Linguistic Ethics Entropy (LEE) index from the perspective of philosophy of technology. This indicator comprehensively considers the ethical performance of machine translation in multiple dimensions such as linguistic symbols, semantics, and culture, and evaluates the degree of compliance with linguistic ethical norms of machine translation by quantitative means. In the context of the development of the EU AI Act, it is recommended to set an access standard of LEE $\leqslant$ 0.23 (Wang, 2024). The setting of this standard can help guide the development of machine translation technology towards compliance with ethical norms at the macro level, guarantee the accuracy, integrity and fairness of language in cross-cultural communication, and promote equal dialogue and mutual understanding between different cultures.

6.3. Interdisciplinary Breakthroughs

Linguistic ethics, as an emerging interdisciplinary discipline, should be actively incorporated into the study of technological object ontology. Taking Derrida's concept of "delayed variation"

as an example, it emphasises the continuous generation and difference of linguistic meaning, which is intrinsically related to the dynamic generation of semantics and the handling of cross-cultural differences in machine translation. Encoding the concept of "delayed variation" into Transformer's positional embedding algorithm can provide new technical ideas for machine translation . Through this interdisciplinary integration, not only can we expand the research scope of language ethics, but also provide theoretical support for the innovation and development of machine translation technology, realise the in-depth cooperation and collaborative innovation among multidisciplinary disciplines, such as linguistics, philosophy, and computer science, and promote the qualitative leap of machine translation technology within the ethical framework. to achieve a qualitative leap within the ethical framework.

Author Contributions

Conceptualization, Z. L. and W. S.; methodology, Z. L. and W. S.; software, Z. L. and W. S.; validation, Z. L. and W. S.; formal analysis, Z. L. and W. S.; investigation, Z. L. and W. S.; resources, Z. L. and W. S.; data curation, Z. L. and W. S.; writing—original draft preparation, Z. L. and W. S.; writing—review and editing, Z. L. and W. S.; visualization, Z. L. and W. S.; supervision, Z. L. and W. S.; project administration, Z. L. and W. S.; funding acquisition, Z. L. and W. S. All authors have read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of Interest:

The author declares no conflict of interest.

Reference

Gao, Y., & Cui, Q. (2024). Challenges and countermeasures of translation system construction in generative AI era. Shanghai Translation, (3), 36-42.

Hu, K., & Li, X. (2023). The development of translation research in the context of large language modelling: Problems and prospects. Chinese Translation, 44(6), 64-73.

Hu, Z., & Tang, Q. (2022). A critical review of Western translation ethics. China Translation, 43(6), 22-29.

Huang, Y. (2025). Building a translation team to meet the needs of the times. Foreign Language Teaching and Research, 57(2), 281-289.

Li, H., & Chen, H. (2020). Translation ethics research in the age of technologisation: Challenges and expansion. Journal of Northeastern University (Social Science Edition), 22(1), 112-119.

Li, Z. (2022). Meta-translation ethics: The cornerstone of translation ethics research. Shanghai Translation, (3), 14-20.

Li, Z. (2023). Returning to ethics: The future road of translation ethics research. Journal of Zhejiang Gongshang University, (1), 24-32.

Liu, C., & Kong, Y. (2023). The essence of ethics in translation technology and its basic direction. Journal of Foreign Languages, (5), 79–85.

Shen, H., Wang, X., & Luo, H. (2025). Machines have machine understanding: Philosophical reflections on the problem of understanding in machine translation. Natural Dialectics Letters, 47(3), 35-41.

Sun, J. (2024). Study on ethical misconduct of machine translation in the context of "Artificial Intelligence + Education". Overseas English, (14), 34-37.

Wang, H. (2024). Research on ethical challenges and response strategies of machine translation. Journal of Qiqihar Higher Teacher Training College, (5), 57-60.

Wang, H., & Liu, W. (2022). Research on translation data ethics in the era of big data: Concepts, issues and suggestions. Shanghai Translation, (2), 12-17, 64.

Wang, H., Cui, Q., & Liu, W. (2024). Research on the ethical risk of generative AI translation technology in legal translation. China Translation, 45(3), 110-117.

Wu, J. (2021). Rethinking and reconstructing translation ethics research. China Translation, 42(5), 5-14.

Wu, M., & Chen, H. (2023). Ethical issues of machine translation in the era of artificial intelligence. Journal of Foreign Languages, (6), 13-18.

Xiao, Z., & Wang, S. (2023). Ethical research on the application of translation technology in the era of artificial intelligence: Concepts, problems and countermeasures. Foreign Language E-learning, (5), 106-112.

Xiong, X., & Wang, Y. (2025). Ethical reflections on machine translation under the theory of "Eight Preparations". Journal of Yichun College, 47(1), 91-97.

Xu, X., Zhai, F., & Yang, L. (2024). An inquiry into the contingency and initiative of constructing the ethical system of machine translation under the perspective of translation ethics. Foreign Trade, (11), 103-105.

Yang, L. (2025). Research on language translation and management of multilingual archives. Archival Practice, (3), 171-173.

Yang, L., & Yang, Y. (2023). Reconceptualising translators' responsibilities in the age of digital technology from the perspective of translation ethics. New Horizons, (2), 145-147.

Yin, M. (2023). A review and prospect of research on the interactive longitudinal nature of translation and ethics: A review of The Lauterich Handbook of Translation and Ethics. Foreign Languages, 46(1), 117-121.

Zhang, L., Zhu, H., & Huang, D. (2024). Sign language translation based on deep learning: Past, present and future. Journal of Software, 35(8), 2938-2963.

Zhang, R., & Zhao, S. (2023). Exploring the ethical load of translation technology from the perspective of actor network theory. Foreign Languages, 46(6), 104-112.

Zhang, X., & Shao, L. (2021). Fundamental principles of constructing trustworthy machine translation systems: A viewpoint based on engineering ethics. Foreign Language, 37(1), 1-8.

Zhang, Y., & Zhang, F. (2024). Study on the outreach translation of marine power strategy from the perspective of ecological translation ethics: Taking Xi Jinping on the Governance of the Country (English Edition) as an example. Journal of Ocean University of China (Social Science Edition), (3), 121-128.

Zhu, J. (2023). Translation studies from the perspective of Marxist philosophy: Current status and prospects. Journal of Foreign Languages, (5), 86-92.