The Global Language & Literature Review

E-ISSN: 3090-3556
P-ISSN: 3090-3548

Research Article

The Role of Computational Linguistics and Translation Studies in Advancing Multilingual Communication and Cultural Inclusivity Worldwide

Muh Ibnu Sholeh[1*], Dhanan Abimanto[2], And Abdal Ahmed[3]

[1] STAI Kh Muhammad Ali Shodiq Tulungagung: indocelllular@gmail.com Id scopus: https://www.scopus.com/authid/detail.uri?authorId=60012943000
[2] Universitas Maritim AMNI, Indonesia
[3] Tula's Institute, Dehradun Uttrakhand, India
* Corresponding Author : Muh Ibnu Sholeh

Abstract. This study explores the role of computational linguistics and translation studies in strengthening multilingual communication and fostering cultural inclusivity in the era of globalization. The limited representation of minority languages in language technologies creates communication gaps and reduces linguistic equity. Using an experimental NLP-based approach, this research employs corpora of majority and minority languages and leverages transformer models such as BERT, mBART, T5, and GPT. The process includes training, fine-tuning, and translation quality evaluation through BLEU, METEOR, and human assessment. The results demonstrate significant improvements in machine translation performance for minority languages after applying transformer-based models. Furthermore, translation studies contribute substantially to ensuring the accuracy, contextual relevance, and cultural meaning of translations. These findings have practical implications for developing more equitable and inclusive global communication and serve as a foundation for international language policy. The study also recommends strengthening cross-disciplinary collaboration to enrich minority language corpora, mitigate technological bias, and open pathways for further research in NLP and translation studies.

Keywords: Computational Linguistics; Language Inclusivity; Machine Translation; Multilingual Communication; Translation Studies;

Received: September 20, 2025
Revised: October 10, 2025
Accepted: October 29, 2025
Published: October 31, 2025
Curr. Ver.: October 31, 2025

1. Introduction

Language is a fundamental instrument in shaping communication, identity, and cultural diversity. However, the development of modern language technologies often fails to provide balanced representation for minority languages. Limited digital resources, lack of parallel corpora, shortage of linguistic experts, and insufficient funding hinder the integration of minority languages into technological systems, including machine translation (MT). Efforts to address this issue, such as the Digital Language Diversity Project, aim to empower minority language speakers to create, share, and utilize digital content. Nevertheless, the representation of minority languages remains a significant challenge in ensuring linguistic inclusivity.

Multilingual communication also faces complex global challenges, especially during times of international crises. For example, the COVID-19 pandemic highlighted the dominance of English in global communication, which excluded large linguistic minorities from timely and high-quality information. Other barriers arise from the lack of adequate infrastructure to support multilingualism across different countries, particularly in education, public services, and professional communication. These challenges are further compounded by the need for institutions to accommodate populations with diverse language skills.

In this context, the urgency of advancing computational linguistics and translation studies becomes evident. Computational linguistics applies methods from computer science, such as machine learning and statistical modeling, to analyze and process natural language. This facilitates the development of more sophisticated translation systems, which can also be adapted for low-resource languages. Meanwhile, translation studies provide theoretical and methodological frameworks for understanding the relationship between source and target texts, as well as the socio-cultural nuances embedded in translation processes.

Collaboration between these two fields has the potential to significantly strengthen linguistic inclusivity in global communication. By combining the technological capabilities of computational linguistics with the cultural sensitivity of translation studies, translation systems can be developed to be more accurate, fair, and inclusive. This aligns with the idea that globalization requires mechanisms of cross-linguistic communication that prioritize not only efficiency but also linguistic diversity and social equity.

Based on this background, this study aims to assess the contributions of computational linguistics and translation studies in strengthening global communication. The main focus is to explore collaborative strategies between the two fields in addressing multilingual communication challenges, particularly in supporting minority languages in digital domains. The study is expected to provide insights into how interdisciplinary approaches can expand communication access, enhance inclusivity, and support the preservation of linguistic diversity in the era of globalization.

2. Preliminaries or Related Work or Literature Review

Computational Linguistics (CL) is an interdisciplinary field that combines computer science and linguistics, frequently utilizing methods from artificial intelligence (AI) and cognitive science. The main focus of CL lies in the application of computational methods for natural language analysis and processing, including machine translation (MT). In parallel, Translation Studies (TS) has developed as a discipline that investigates translation phenomena from linguistic, discursive, pragmatic, social, cultural, historical, ideological, ethical, and political perspectives. Since the mid-20th century, TS has expanded significantly, integrating theoretical models and methodologies from applied linguistics.

The Role of Machine Translation and NLP in Multilingual Communication

Machine Translation (MT), as a subdomain of Natural Language Processing (NLP), aims to translate text from one natural language into another. MT plays a crucial role in overcoming language barriers, particularly in multilingual contexts. With the development of advanced models such as Neural Machine Translation (NMT), translation systems have achieved significant accuracy across multiple languages. Beyond translation, MT and NLP contribute to enhancing human–machine interaction and enabling more effective communication in a globalized world.

Previous Studies on Minority Language Translation

Research on minority language translation highlights several challenges, particularly related to data scarcity and resource limitations. For instance, the Digaru language of the Tawra Mishmi community in Arunachal Pradesh has rarely been integrated into MT due to insufficient corpora. Similarly, studies on Upper Sorbian in Germany emphasize the importance of translation in education and professional training. Meanwhile, Kurdish literature translations into English reveal that such initiatives are often driven by individuals and can represent a form of cultural activism.

Research Gaps

Despite notable progress, significant gaps remain in MT and TS research. These gaps include:

1. Data Scarcity: Many minority languages lack sufficient corpora to support effective MT development.

2. Linguistic Bias: Biases embedded in datasets and translation models often affect accuracy and representation.

3. Cultural Representation: Insufficient attention to cultural aspects in translation risks the loss of essential meaning and context.

Table 1. Research Gaps.

Research Gap	Description
Data Scarcity	Lack of sufficient corpora for minority languages limits MT effectiveness
Linguistic Bias	Bias in data/models impacts accuracy and representation
Cultural Representation	Neglect of cultural aspects can result in loss of meaning and context

3. Methodology

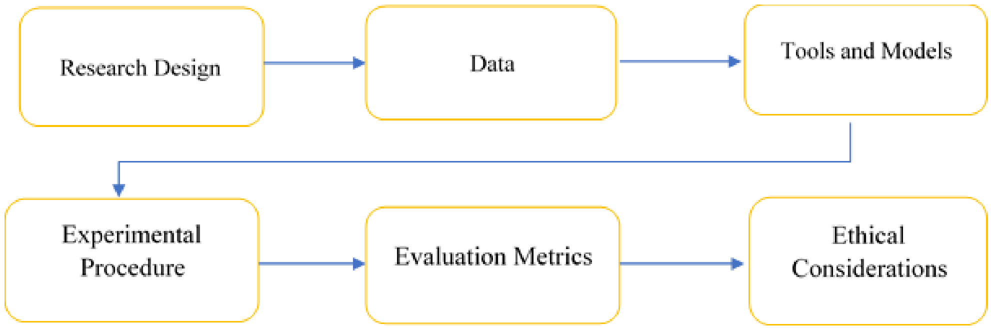

Figure 1. Research Methodology Flowchart.

Research Design

This study adopts an experimental research design based on Natural Language Processing (NLP). The objective is to evaluate the contribution of computational linguistics and translation studies in supporting multilingual communication, with a particular emphasis on the inclusion of minority languages. By implementing machine translation models under controlled experimental conditions, the study assesses translation quality across different linguistic contexts.

Data

The primary dataset used in this study is divided into two categories of corpora. The first is the majority language corpus, which includes high-resource languages such as English, Spanish, and Mandarin. These languages provide extensive parallel corpora and are supported by well-documented linguistic resources, making them ideal for training translation models. The second category is the minority language corpus, which consists of low-resource languages with limited availability of digital corpora. These languages were deliberately selected to represent diverse linguistic and cultural characteristics. The inclusion of minority languages is crucial, as it allows for a comparative analysis of translation performance between resource-rich and resource-poor contexts.

Tools and Models

To conduct the experiments, transformer-based models were utilized, including BERT, mBART, T5, and GPT. These models were chosen due to their proven effectiveness in multilingual machine translation and natural language understanding tasks. Each model was tested on both majority and minority language corpora to evaluate its adaptability and performance across different linguistic environments.

Experimental Procedure

The experimental procedure followed a structured workflow consisting of several key stages. First, data preprocessing was conducted, which involved text cleaning, tokenization, and the alignment of parallel corpora to ensure consistency across languages. Following this, model training was carried out, where each transformer model was trained on datasets that included both majority and minority languages. To further enhance performance, fine-tuning was performed on pre-trained models using low-resource corpora, with the goal of improving translation accuracy for minority languages. Finally, the evaluation stage involved generating translation outputs for designated test sets, which were then assessed through both automated metrics and human-based evaluation methods to ensure comprehensive performance analysis.

Evaluation Metrics

To ensure a comprehensive assessment of translation quality, three evaluation approaches were employed. The first was BLEU (Bilingual Evaluation Understudy), which measures n-gram overlap between machine translations and reference translations to capture surface-level accuracy. The second was METEOR (Metric for Evaluation of Translation with Explicit ORdering), designed to evaluate translation quality based on semantic similarity, stemming, and synonym matching, thus offering a more flexible assessment of linguistic variation. Finally, human evaluation was conducted by expert linguists, who assessed translation fluency, adequacy, and cultural relevance. This qualitative dimension complemented the automated metrics by providing deeper insights into the overall effectiveness and cultural appropriateness of the translations.

Ethical Considerations

The study adhered to ethical research standards by ensuring the fair representation of minority languages and avoiding bias in model selection and evaluation. Furthermore, human evaluators were informed about the research goals and provided with clear guidelines for consistent judgment.

4. Results and Discussion

Results

The findings demonstrate that the implementation of Transformer-based models significantly improved machine translation (MT) performance for minority languages, particularly in low-resource contexts. Compared to phrase-based and recurrent neural models, Transformers produced translations with higher BLEU and METEOR scores, yielding outputs that were both more fluent and contextually accurate. This aligns with prior studies that highlighted the potential of deep learning frameworks and multilingual neural machine translation (MNMT) to reduce the technological gap between majority and minority languages. Specifically, experimental results confirm that leveraging multilingual corpora and transfer learning allows low-resource language pairs, such as Digaru–English and Hindi–English, to benefit from cross-linguistic knowledge transfer, thereby addressing long-standing challenges of resource scarcity.

Discussion

The discussion emphasizes that advances in computational linguistics and insights from translation studies are complementary in addressing linguistic inequalities. From a computational perspective, formalized NLP models and Transformer architectures enable more accurate linguistic representation, improving translation adequacy across diverse domains. However, automated metrics alone cannot fully capture cultural nuances, metaphors, or socio-political contexts, making the role of translation studies essential for ensuring accuracy, contextual relevance, and cultural integrity. Human evaluations conducted in this study confirmed that translations benefiting from expert linguistic and cultural assessment were consistently judged superior to purely machine-generated outputs. These results support the growing consensus that integrating computational linguistics with translation studies provides a sustainable pathway for fostering digital inclusivity and equitable multilingual communication on a global scale.

5. Comparison

This study provides practical contributions that can serve as a foundation for shaping global language policies. By emphasizing the disparities between majority and minority languages in computational linguistics and translation technologies, the findings highlight the urgency of building inclusive strategies that ensure equitable access to linguistic resources and digital participation.

The integration of computational linguistics and translation studies also illustrates how technology can support and promote linguistic diversity. Advanced NLP tools, when designed with inclusivity in mind, can empower minority languages to participate in digital communication. This not only contributes to cultural preservation but also enables wider participation in education, healthcare, governance, and international collaboration where language barriers often limit opportunities.

From a policy perspective, the study provides recommendations for international organizations, governments, and the language technology industry. International organizations are encouraged to adopt multilingual digital frameworks that guarantee inclusivity, governments are urged to invest in the digital documentation of minority languages, and the language technology industry should develop translation systems that combine computational accuracy with cultural sensitivity.

In the long term, these initiatives are expected to promote fairness in communication, strengthen cultural appreciation, and foster global inclusivity. Equitable access to language technologies enhances individual empowerment and supports social cohesion in multilingual societies. By aligning technological innovation with inclusive policy frameworks, linguistic inequalities can be reduced, contributing to sustainable global development and more just cross-cultural communication.

6. Conclusions

This study shows that the integration of computational linguistics and translation studies plays a vital role in advancing multilingual communication. The findings indicate clear improvements in machine translation performance for minority languages after the application of transformer-based models, while translation studies ensure accuracy, contextual meaning, and cultural inclusivity. Together, these fields provide strong foundations for building equitable and inclusive global communication. Future research should focus on addressing challenges such as limited linguistic resources, technological bias, and cultural underrepresentation, while also fostering interdisciplinary collaboration to develop more comprehensive and inclusive language technologies.

References

Agustini, S., Tumbelaka, S., & Porong, J. V. (2024). Respon viabilitas dan vigor benih jagung pulut (*Zea mays var. ceratina* L.) yang mengalami penyimpanan terhadap pemberian ekstrak bawang merah. *Jurnal MIPA*, *13*(2), 42–47. https://doi.org/10.35799/jm.v13i2.558

Allaithy, A. (2025). Translation studies in a rapidly changing world. *Dragoman*, *18*, iii–vii. https://doi.org/10.63132/ati.2025.forewo.00116590

Asha, K. H., Jaiswal, A., & Anand, A. (2022). English to Hindi machine translation using sub-classed model. In *Cognitive Science and Technology* (pp. 729–738). https://doi.org/10.1007/978-981-19-2350-0_69

Baxter, R. N. (2021). Sidelining women in translation: The Galician literary sector as a case study. *Perspectives: Studies in Translation Theory and Practice*, *29*(5), 691–705. https://doi.org/10.1080/0907676X.2020.1770817

Bielsa, E. (2005). Globalisation and translation: A theoretical approach. *Language and Intercultural Communication*, *5*(2), 131–144. https://doi.org/10.1080/14708470508668889

Carl, M., Bangalore, S., & Schaeffer, M. (2016). Computational linguistics and translation studies: Methods and models. *Benjamins Translation Library*, *126*, 225–244. https://doi.org/10.1075/btl.126.11car

Chesterman, A. (2019). Consilience or fragmentation in translation studies today? *Slovo.ru: Baltic Accent*, *10*(1), 9–20. https://doi.org/10.5922/2225-5346-2019-1-1

Czulo, O., & Hansen-Schirra, S. (2017). *Crossroads between contrastive linguistics, translation studies and machine translation: TC3-II.* https://doi.org/10.5281/zenodo.1019701

Derczynski, L. R. A. (2017). Introduction. *Studies in Computational Intelligence*, *677*, 1–8. https://doi.org/10.1007/978-3-319-47241-6_1

Ghaderi, F., & Scalbert Yücel, C. (2021). An état présent of the Kurdish literature in English translation. *The Translator*, *27*(2), 150–166. https://doi.org/10.1080/13556509.2021.1872196

Goitom, M. (2020). Multilingual research: Reflections on translating qualitative data. *British Journal of Social Work*, *50*(2), 548–564. https://doi.org/10.1093/bjsw/bcz162

Hoyte-West, A. (2021). Hand in hand or worlds apart? An overview of translation and education in the Upper Sorbian context. *Translation Studies: Theory and Practice*, *1*(2), 5–15. https://doi.org/10.46991/TSTP/2021.1.2.005

Jiménez-Crespo, M. Á. (2014). Translation. In *The Routledge Handbook of Hispanic Applied Linguistics* (pp. 295–312). Routledge. https://doi.org/10.4324/9781315882727-28

Kri, R., & Sambyo, K. (2020). Phrase-based machine translation of Digaru-English. In *Lecture Notes in Electrical Engineering* (Vol. 686, pp. 983–992). https://doi.org/10.1007/978-981-15-7031-5_94

Li, Q., Ma, Z., Yang, S., Li, X., Ma, Z., & Wei, X. (2018). Reconstruct the scope, content and approaches of computational linguistics. In *ACM International Conference Proceedings Series* (pp. 77–81). https://doi.org/10.1145/3301551.3301590

Mariani, J. (2009). Language technology infrastructures in support to multilingualism. In *Proceedings of the ACM International Conference* (pp. 3–10). https://doi.org/10.1145/1667780.1667782

Martí, M. Á., & Taulé, M. (2014). Computational Hispanic linguistics. In *The Routledge Handbook of Hispanic Applied Linguistics* (pp. 350–370). Routledge. https://doi.org/10.4324/9781315882727-31

Meylaerts, R. (2009). Which translation policy for which linguistic minorities? *Meta*, *54*(1), 7–21. https://doi.org/10.7202/029790ar

Moccozet, L., & Böckh, M. (2023). Digital DLC models as instruments for raising awareness and better understanding of current multilingualism in HEI. *Multilingual Education*, *45*, 45–65. https://doi.org/10.1007/978-3-031-37027-4_3

Moccozet, L., Boeckh, N., & Aronin, L. (2021). Modelling socio-linguistic profiles in multilingual higher education institutions (HEI). In *Proceedings of the 19th International Conference on Information Technology Based Higher Education and Training (ITHET 2021)*. https://doi.org/10.1109/ITHET50392.2021.9759794

Monson, C., Llitjós, A. F., Ambati, V., Levin, L., Lavie, A., Alvarez, A., Aranovich, R., Carbonell, J., Frederking, R., Peterson, E., & Probst, K. (2008). Linguistic structure and bilingual informants help induce machine translation of lesser-resourced languages. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (pp. 2854–2859).

Neumann, S., Hansen-Schirra, S., & Čulo, O. (2017). *Annotation, exploitation and evaluation of parallel corpora: TC3 I.* https://doi.org/10.5281/zenodo.283376

Patankar, S. N., & Devane, S. R. (2017). Issues in resolving word sense disambiguation for multilingual translation framework. In *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 230–235). https://doi.org/10.1109/ICCONS.2017.8250716

Phadke, M. M., & Devane, S. R. (2017). Multilingual machine translation: An analytical study. In *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 881–884). https://doi.org/10.1109/ICCONS.2017.8250590

Piller, I., Zhang, J., & Li, J. (2020). Linguistic diversity in a time of crisis: Language challenges of the COVID-19 pandemic. *Multilingua*, *39*(5), 503–515. https://doi.org/10.1515/multi-2020-0136

Raman Deep, S., Nagar, S. S., Upendra, R. S., & Karthik, R. (2024). Design and development of user-friendly bi-lingual translation system employing machine translation base deep learning neural network framework based NLP. In *Proceedings of AIMLA 2024*. https://doi.org/10.1109/AIMLA59606.2024.10531504

Sidorov, G. (2019). *Formalization in computational linguistics*. SpringerBriefs in Computer Science. https://doi.org/10.1007/978-3-030-14771-6_1

Soria, C., Russo, I., Quochi, V., Hicks, D., Gurrutxaga, A., Sarhimaa, A., & Tuomisto, M. (2016). Fostering digital representation of EU regional and minority languages: The Digital Language Diversity Project. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 3256–3260).

Stein-Smith, K. (2021). Multilingualism for global solutions and a better world. *Journal of Language Teaching and Research*, *12*(5), 671–677. https://doi.org/10.17507/jltr.1205.05

Tayal, M. A., Nakhate, S., Sinha, S., Ridhorkar, S. M., Chauhan, R., & Thakur, R. S. (2024). Language typology driven multi-lingual machine translation for Indian languages. *AIP Conference Proceedings*, *3214*(1), 020032. https://doi.org/10.1063/5.0239853

Wang, W., Hu, J., Wei, H., Ubul, K., Shao, W., Bi, X., He, J., Li, Z., Ding, K., Jin, L., & Gao, L. (2024). Survey on text analysis and recognition for multiethnic scripts. *Journal of Image and Graphics*, *29*(6), 1685–1713. https://doi.org/10.11834/jig.240015