Centre for International Governance Innovation

Digital Policy Hub

Digital Policy Hub – Working Paper

Language Translation in the AI-Augmented Era

Anna Legault

Fall term – 2025–2026 cohort

About the Hub

The Digital Policy Hub at CIGI is a collaborative space for emerging scholars and innovative thinkers from the social, natural and applied sciences. It provides opportunities for undergraduate and graduate students and post-doctoral and visiting fellows to share and develop research on the rapid evolution and governance of transformative technologies. The Hub is founded on transdisciplinary approaches that seek to increase understanding of the socio-economic and technological impacts of digitalization and improve the quality and relevance of related research. Core research areas include data, economy and society; artificial intelligence; outer space; digitalization, security and democracy; and the environment and natural resources.

The Digital Policy Hub working papers are the product of research related to the Hub's identified themes prepared by participants during their fellowship.

Partners

Thank you to Mitacs for its partnership and support of Digital Policy Hub fellows through the Accelerate program. We would also like to acknowledge the many universities, governments and private sector partners for their involvement allowing CIGI to offer this holistic research environment.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

About the Author

Anna Legault is a Digital Policy Hub master's fellow pursuing an M.A. in digital humanities at McGill University. With a background in linguistics and translation for the United Nations and Canadian government, she is a practitioner-academic of interdisciplinary language work. As part of the Digital Policy Hub fellowship, she will be examining the role of emerging language translation technologies in geopolitics.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
cigionline.org

Policy Memo

Language Translation in the AI-Augmented Era

Anna Legault

Bottom Line Up Front

Cross-linguistic dialogue is a critical aspect of global collaboration and is upheld by language translation practices. These practices are evolving. Emerging large language model (LLM)-powered technologies for language translation are integrating into institutional settings and professional norms. The influx of artificial intelligence (AI) into language tasks, mixed with the culturally nuanced nature of language surrounding geopolitical concepts, makes understanding the limits of LLMs especially relevant to institutions that rely on translation and host political dialogue. Language operations policies must be recodified in light of the AI-augmented era. Technology and humans engage in different processes of meaning-making; they also have different relationships to liability. An updated delimitation of roles in human-technological partnership can benefit from the strengths of each actor while mitigating the risk of automation. Especially where the stakes with regard to accuracy are high, responsible AI-augmented translation is human-centred and accounts for culturally specific mediation between linguistic groups.

Key Points

- Translation is not a neutral channel for information to pass through. The linguistic intermediary is an active contributor to the co-construction of meaning and is responsible for countless micro-decisions that influence overall outcomes. Language encodes cultural ontologies, thus, translation requires a subjective mediation on how to bridge cultural knowledge systems.
- Humans and LLMs diverge in their paths to meaning-making. LLM architectures form networks of association between concepts through vectorization and word embedding, enabling computation. Human languages contain category systems that vary across cultures.
- The development of deep neural networks for multilingual LLMs has resulted in progress for the accuracy of multilingual computation, but limits persist around cultural nuance.
- When an LLM generates any part of language used by an institution, it is an agentive actor in that context. The extent of this agency should correspond to the potential outcomes of the dialogue it creates.

Recommendations

- **Redesign institutional language operations with increased procedural transparency:** Structuring dialogue with the linguistic mediator positioned as part of the interaction rather than in the margins increases transparency around the chosen level of automation, as well as the choices being made to bridge languages. Bringing language processes into the systems design of institutional operations responsibly accounts for the influence of the intermediary.
- **Prioritize a human-centred augmented translation framework:** Leverage technological strengths to empower human actors rather than replacing the full translation process. Especially where meaning systems between languages diverge the most and where language has political implications, human language professionals must exercise trained judgement on relating systems of meaning while potentially being supported by the computational power of automated subtasks.
- **Delimit use-cases for automated translation based on importance:** Codifying a threshold beyond which oversight is needed and liability can be assigned to the appropriate entity would contribute to informed use of technology and its uniform deployment across proceedings.

Introduction

The attempt to automate translation between the world's natural languages has been a priority through many waves of the creation of artificial intelligence (AI). From early algorithms that programmed "rule-based translation" according to grammar, to statistical machine translation that calculated a probable output based on parallel bilingual corpora, the performance of translation technology has been steadily on the rise (Kenny 2022, 35). The current era of deep neural networks brings with it the capacity for multilingual large language models (LLMs) to encode many languages and to translate between languages that were not originally paired in the training data (Chen and Zhang 2024). This shift expands the horizons of the technology as well as the pairings of languages between which translation can occur. Concurrently, professional translators and interpreters have been utilizing digital tools in their work previous to the existence of multilingual LLMs. Translation memory systems, quality estimation and automated content enrichment are all examples of technological supports to the translation process that have contributed to a norm of human-technological collaboration (O'Brien 2023). The current "technological turn" in the language professions is being documented with strategies such as "computer-assisted" and "remote" interpreting (Fantinuoli 2018, 4). The newfound performance of multilingual LLMs means automated translation technologies are integrating into professional practice and institutional use alike; this shift has been described as the "augmented paradigm" (Fantinuoli and Dastyar 2022).

As AI and the language professions continue to co-evolve, questions arise around what can be responsibly automated. Translators and interpreters have always been the intermediaries of our world's most high-stakes conversations in the international milieu. The European Commission's Directorate-General for Translation employs several thousand internal and external translators for the EU's institutions,[1] while translation booths line the uppermost level of the United Nations General Assembly. Facilitated multilingual communication is a key component of global collaboration, and international institutions are now tasked with (re)codifying language operations standards for the augmented era. Different approaches to the task are evident: the United Nations Office at Geneva, for example, translates its official information online into the languages of its member states in accordance with policies of multilingualism and access, currently with the disclaimer "some parts have been translated using machine translation software and accuracy cannot be guaranteed; revision is pending, subject to availability of resources."[2] As institutions grapple with the opportunity and risk of emerging translation technologies, understanding the limits of LLMs is key in contexts that rely on translation and host political dialogue. This working paper analyzes recent published conference proceedings on multilingual LLM progress through the lens of sociolinguistics and translation studies. Generalized approaches to meaning-making by human and technological actors are compared. Interrogating notions of equivalence, agency and human parity inform recommendations for augmented translation policy in an institutional context.

1 See https://commission.europa.eu/about/departments-and-executive-agencies/translation_en.

2 See www.ungeneva.org/en/multilingualism-disclaimer.

Constructing Meaning

Translation requires the transfer of meaning from one language to another, yet meaning is constructed differently across languages. Thus, the intermediary must communicate something that is as close to the original idea as possible while using the words available in the target language. Functionalist translation theory posits that exact one-to-one transfer is impossible; equivalence as a concept is treated more like a goal (Nord 2010). The fallacy of equivalence is especially evident as concepts become more abstract, metaphoric or culturally specific. As such, the translator is not a neutral channel but a cultural mediator and an active agent in the co-construction of meaning (Wadensjö 1999). Should any part of that process be automated, an understanding of how meaning is built by the technological actor is relevant to crafting guidelines for its responsible use.

Professional translators and interpreters are extensively trained in the situated relationship between two languages in accordance with best practices born from the canon of translation theory (Pattison and Cragie 2022). In contrast, LLMs are trained on vast data sets of available digital information and encode language through word embedding. The following is a generalized comparison of routes to meaning-making taken by both of these approaches.

LLM Architectures: Vectorization and Word Embedding

The first step in neural machine translation, converting from language to numerical values that can be calculated by the LLM, is itself a translation. Word embedding assigns fixed vectors to all words in a language such that concepts that are similar to one another are closer together and dissimilar concepts are farther apart. For example, the English word for “table” might have a very similar numerical representation to that of “chair” because those two words would have often co-occurred in training data, resulting in their numerical association. Graphing this along an x, y axis would place them together, whereas a word like “consequential” would appear farther away. LLM architectures map languages spatially using vectors (Chen and Zhang 2024). They do so not only in two-dimensional or three-dimensional space, but also to the nth dimension, beyond what is easy for humans to visualize but feasible for the neural network to compute. Multilingual LLMs convert between languages that have been mapped in this way, associating geometric and numerical similarities across languages before ultimately creating an output in the target language.

An analysis of recent published proceedings from the International Joint Conference on Neural Networks (IJCNN) reveals progress toward accurate multilingual transfer between vectorized word embeddings, with persistent limits: one study concludes that the “systems encounter difficulties in thoroughly unravelling the culturally nuanced and intricate semantic variations inherent in multilingual computation” (Zhang, Mao and Cambria 2024). This particular conclusion arose as a result of differing social meaning between so-called equivalent emotion terms in English and Japanese. In this example, cultural response to the words was part of meaning-making beyond the lexical association of terms. An emergent theme through the conference papers was an evaluation of “zero-shot machine translation” (Chen and Zhang 2024). Zero-shot refers to translation between languages that were not directly paired in LLM training,

relying instead on the model's ability to generalize shared linguistic representations learned from many mapped languages. The idea underpinning this method is that all languages are grounded in a perception of the human experience. Some experiences and their corresponding concepts may be highly similar across cultures, enabling common numerical representation and easier translation. In one early example of this cross-linguistic embedding, Google contributors graphed the vectors for words between one and five, in English and in Spanish respectively. While the word-embedding process had occurred within monolingual training data, the numbers appeared at highly similar spatial locations (Mikolov, Le and Sutskever 2013). This is because the concept of "five" had a very similar linguistic relationship to the words for other numbers in both English and in Spanish. The relationships between words for numbers are broadly grounded in the real world and present across languages. As concepts become more complex, so too does associating vectors across linguistic divides. Emerging strategies such as "cross-lingual word embedding," where the model initially learns the numeric representations of words "in a joint vector space" (i.e., from data in both languages), has shown promise in addressing cultural norms (Kumar and Sahula 2022).

Word embedding represents each word as a numerical vector. These vectors demonstrate the proximity of that word to all other words in the language according to patterns in the training data. The amount of training data available varies by language, influencing the quality of mapping. Current methods of vectorization broaden the potential for translation across diverse language pairs, and have achieved fair performance outcomes to date. However, word embedding associates concepts without situated social context, potentially resulting in shallow or misleading translations (depending on the language and model).

Culture and Categorization

Multilingual LLMs map the relational proximity between concepts based on occurrence in their training data; this can be analyzed as a kind of web of associative bonds. Beyond association, schools of thought in sociolinguistics advance the notion that human languages encode the relationships between their words based on culturally specific ontologies (Ji, Zhang and Nisbett 2004; Katan 2009). Category systems exist in language to "interact with our environment without becoming overwhelmed by its complexity" (Wen and Fu 2021, 173) and have been thoroughly studied from a cross-cultural perspective. Seminal work such as George Lakoff's *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (1987) appealed for a shift away from objectivism and generalization, in favour of experiential and embodied cognition. Colour terms have been a frequent site for this analysis on the basis that all human languages have developed with access to the same spectrum of light, yet colour nomenclatures differ across language families. In one example, speakers of Russian consistently selected different boundaries around what they considered "blue" compared to English speakers when presented with the same colour chips (Bimler and Uusküla 2017). In another study, Polish speakers categorically chose a term for "blue" based on whether the connotation was positive or negative (Stanulewicz 2010). Universalist versus relativist debates prevail; however, evidence of linguistic relativity suggests that culture, codified by language, has some role in perception and expression. Human category systems (such as whether concepts relate to each other hierarchically, horizontally or by radial network) operate beyond the simple association of related terms, influenced by the cultural

context and the conceptual priorities of a given language community. Whether the encoding of how words relate to one another is done by training an LLM, or by humans speaking in their cultural context, the conceptual proximity between words is an artifact of social values in the learning environment. Different mappings will result in different linguistic outcomes.

If categorization and knowledge systems differ among languages, the act of translation necessarily includes cross-cultural bridging and mediation. If the proximity between words is an artifact of social values, translation entails the mapping of one language's values onto another language. This specifically pertains to international institutions since there is an additional complexity and negotiation of power surrounding political language. Translation studies have mapped the way translators skillfully navigate geopolitical differences in multilingual communication, though evidence demonstrates that meaning shift can still occur during this process: James Chonglong Gu (2019) transcribes side-by-side glosses of the Chinese premier fielding interpreter-mediated questions from international journalists and finds salient ideological shifts (e.g., "Taiwan" versus "Taiwan local region"; 9). Morven Beaton (2007) examines a corpus of European Parliament transcripts, finding pre-existing power balances within the institution to be reinforced through the interpretation process. When worldview, linguistic categories and geopolitical interests diverge, the complexity of translation increases alongside the level of influence enacted by the intermediary. Responsible integration of any emerging language translation technologies must account for the geopolitical consequences of those technologies' linguistic choices.

Agency in Augmented Translation

Lisa McEntee-Atalianis and Rachelle Vessey (2024) use corpus linguistics to analyze agentive actors in proceedings at the United Nations. They show that conducting dialogue in an institutional context inherently enacts agency, and that the dialogic process itself is both constrained by the limits of its environment as well as potentially subversive to those limits. An agent is defined here as having "the power to enact, (re)interpret, negotiate or challenge policy" (ibid., 360). Through this lens, multilingual LLMs are agentive actors should they generate any part of language used in institutional operations. This is further supported by a translation studies approach that positions the translator as a co-creator of meaning. This raises questions as to what decision-making power institutions allot to technology. With linguistic decisions occurring as a result of an LLM's subjective word-embedding architecture, the agency vested in that technology may influence a reinforcement or subversion of power relations where the technology is used.

Emerging literature on the augmented paradigm takes inspiration from frameworks of "human centered artificial intelligence" to theorize human-centred augmented translation (O'Brien 2023, 392). It is argued that automation is not the same as loss of control. To conflate the two is described as an "antagonistic dualism" of human versus machine (ibid.). Instead of seeking to replace or surpass all facets of human-based translation, the focus of this framework remains on leveraging the technology

to empower translators by automating parts of the process while maintaining human control. Supporting human creativity and cognitive load is a benefit of the augmented paradigm, potentially empowering translators to become more efficient and accurate through the speed and breadth of linguistic computation. This would, however, require a deployment of technology that is tailored to the complexity of translation and used to support rather than to replace the full process. Further building on the precedent of technologically-supported human translation could look like utilizing LLM performance to suggest options based on dialect, offer historical context or instantaneously search for relevant translation examples. Fully automated translation may be appropriate depending on the stakes, with a degree of human oversight that corresponds to the case, and procedural transparency around the use of technology. With current evidence demonstrating LLM limits around cultural nuance, translation that requires that most mediation between cultural systems suggests a need for the most human involvement.

Liability

Evaluators of AI have often used the term "human parity" as a benchmark when considering the performance of a model; it was even used to advertise LLM-based translation products when neural networks allowed for better performance of metaphoric and idiomatic expressions in the mid-2010s (Kenny 2022, 39). Gauging whether technology is on par with (or surpassing) human capability is useful when analyzing performance on specific tasks. However, when technology is not beholden to the same systems of accountability as humans, human parity is not relevant to liability. This implicates translation in institutions where accuracy has legal implications, errors require transparency and culturally specific equivalence can be subjective. Constructing ideas across languages by building between cultural systems of meaning is a process with ethical decision making involved. Multilingual LLMs certainly surpass human performance on tasks such as recall, search and calculation of many variables, but an augmented paradigm of translation requires that guidelines ensure accountability remains with an actor that can be held accountable. According to Madeleine Clare Elish (2019, 55), "when humans and machines work together, traditional conceptions of control and responsibility will likely need to change in response" especially where "control over an action or function has become distributed across multiple actors (human and nonhuman)." Institutions recodifying language operations must thus explicitly assign where liable professionals are needed (e.g., where dialogue has high-consequence outcomes, or where the translation task has heightened ethical decision making).

Recommendations

- **Redesign institutional language operations with increased procedural transparency:** Instead of relying on a conduit model that assumes translation is neutral, design institutional language operations in a way that accounts for the influence of translation. Structurally bringing language roles from the institutional margins into an equally participatory role increases transparency around the decisions made by the intermediary and any meaning shift that may occur during translation. This redesign must include the disclosure of any level of automation involved in language work. An approach to facilitated dialogue as an interaction

between all actors, with space allotted for a linguistic intermediary to interject, report errors and disclose culturally relevant semantic shift could influence accuracy and monitor issues with automation.

- **Prioritize a human-centred augmented translation framework:** Leverage technological strengths to empower humans actors rather than replacing the full translation process, especially during communication where meaning systems between languages are most divergent and where language has political implications. Current models demonstrate limits around aligning the ontologies that are encoded within languages. This gap is an ideal site for human language professionals to exercise trained judgement on relating systems of meaning while being supported by the computational power of automated subtasks.
- **Delimit use-cases for automated translation based on importance:** Codifying a threshold beyond which human oversight is needed and liability is assigned to the appropriate entity allows processes of accountability to safeguard proceedings. This includes the standardization of when automation is off-limits. A use-case of low importance might allow for language to be fully generated by a multilingual LLM, however, much of the communication present in international institutions has broad global implications. This necessitates specific delimitation of the topics and contexts that require a liable party and a human ethical agent as an intermediary.

Acknowledgements

Thank you to the Digital Policy Hub, under the excellent leadership of Dianna English, for the opportunity to conduct this work. This working paper significantly benefited from guidance by CIGI Senior Fellow Maroussia Lévesque, peer reviewer Daniel Dela Cruz, and my academic supervisor in the Digital Humanities, Dr. Cecily Raynor. I would also like to express my gratitude for inspiration and insights from Caleigh Wong and for mentorship in language work by Deloris Piper.

Works Cited

Beaton, Morven. 2007. “Intertextuality and Ideology in Interpreter-Mediated Communication: The Case of the European Parliament.” Ph.D. dissertation, Heriot-Watt University. www.ros.hw.ac.uk/handle/10399/2028.

Bimler, David, and Mari Uusküla. 2017. “A Similarity-Based Cross-Language Comparison of Basicness and Demarcation of ‘Blue’ Terms.” *Color Research and Application* 42 (3): 362–77. https://doi.org/10.1002/col.22076.

Chen, Xiao and Chirui Zhang. 2024. “Language-Agnostic Zero-Shot Machine Translation with Language-Specific Modeling.” In *Proceedings of the 2024 International Joint Conference on Neural Networks* (IJCNN), 1–8. Yokohama, Japan. https://doi.org/10.1109/IJCNN60899.2024.10649983.

Elish, Madeleine Clare. 2019. “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction.” Engaging Science, Technology, and Society 5 (March): 40–60. https://doi.org/10.17351/ests2019.260.

Fantinuoli, Claudio. 2018. “Interpreting and Technology: The Upcoming Technological Turn.” In *Interpreting and Technology*, edited by Claudio Fantinuoli, 1–12. Berlin: Language Science Press. https://langsci-press.org/catalog/book/209.

Fantinuoli, Claudio and Vorya Dastyar. 2022. “Interpreting and the Emerging Augmented Paradigm.” *Interpreting and Society* 2 (2): 185–94. https://doi.org/10.1177/27523810221111631.

Gu, James Chonglong. 2019. “Interpreters Caught Up in an Ideological Tug-of-War? A CDA and Bakhtinian Analysis of Interpreters’ Ideological Positioning and Alignment at Government Press Conferences.” *Translation and Interpreting Studies* 14 (1): 1–20. https://doi.org/10.1075/tis.00027.gu.

Ji, Li-Jun, Zhiyong Zhang and Richard E. Nisbett. 2004. “Is It Culture or Is It Language? Examination of Language Effects in Cross-Cultural Research on Categorization.” *Journal of Personality and Social Psychology* 87 (1): 57–65. https://doi.org/10.1037/0022-3514.87.1.57.

Katan, David. 2009. “Translation as Intercultural Communication.” In *The Routledge Companion to Translation Studies*, edited by Jeremy Munday, 74–92. London, UK: Routledge. https://doi.org/10.4324/9780203879450-12.

Kenny, Dorothy, ed. 2022. *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*. Berlin, Germany: Language Science Press. https://doi.org/10.5281/zenodo.6653406.

Kumar, Rashi and Vineet Sahula. 2022. “Word Translation Using Cross-Lingual Word Embedding: Case of Sanskrit to Hindi Translation.” In *Proceedings of the 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*, 1–7. Vijayawada, India. https://doi.org/10.1109/AISP53593.2022.9760564.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: University of Chicago Press.

McEntee-Atalianis, Lisa and Rachelle Vessey. 2024. “Using Corpus Linguistics to Investigate Agency and Benign Neglect in Organisational Language Policy and Planning: The United Nations as a Case Study.” *Journal of Multilingual and Multicultural Development* 45 (2): 358–73. https://doi.org/10.1080/01434632.2021.1890753.

Mikolov, Tomas, Quoc V. Le and Ilya Sutskever. 2013. “Exploiting Similarities among Languages for Machine Translation.” arXiv, September 17. https://arxiv.org/pdf/1309.4168.

Nord, Christiane. 2010. “Functionalist Approaches.” In *Handbook of Translation Studies*, edited by Yves Gambier and Luc van Doorslaer, vol. 1, 120–28. Amsterdam, the Netherlands: John Benjamins.

O’Brien, Sharon. 2023. “Human Centered Augmented Translation: Against Antagonistic Dualisms.” *Perspectives: Studies in Translation Theory & Practice* 32 (3): 391–406. https://doi.org/10.1080/0907676X.2023.2247423.

Pattison, Ann and Stella Cragie. 2022. *Translating Change: Enhanced Practical Skills for Translators*. New York, NY: Routledge. https://doi.org/10.4324/9781003136903.

Stanulewicz, Danuta. 2010. “Polish Terms for ‘Blue’ in the Perspective of Vantage Theory.” *Language Sciences* 32 (2): 184–95. https://doi.org/10.1016/j.langsci.2009.10.004.

Wadensjö, Cecilia. 1999. “Interpreters and Other Intermediaries.” In *Interpreting as Interaction*, 35–52. London, UK: Routledge. https://doi.org/10.4324/9781315842318.

Wen, Xu and Zhengling Fu. 2021. “Categorization.” In *The Routledge Handbook of Cognitive Linguistics*, edited by D. Louise Taylor, 173–92. London, UK: Routledge. https://doi.org/10.4324/9781351034708.

Zhang, Xulang, Rui Mao and Erik Cambria. 2024. “Multilingual Emotion Recognition: Discovering the Variations of Lexical Semantics between Languages.” In *Proceedings of the 2024 International Joint Conference on Neural Networks* (IJCNN), 1–9. Yokohama, Japan. https://doi.org/10.1109/IJCNN60899.2024.10651409.