Trans-Babelism Theory: English, AI, and the Transformation of Global Linguistic Order

Hoang-Nam Tran, Tokushima University, Japan

The Southeast Asian Conference on Arts & Humanities 2026
Official Conference Proceedings

Abstract

This paper proposes the Trans-Babelism (TB) theory to interpret the shifting dynamics between English and global multilingualism in the age of AI. For decades, English has served as the modern Tower of Babel symbolizing both unity and inequality in global communication. Yet AI-driven translation, speech recognition, and generative language models are now dissolving linguistic boundaries, creating the possibility of a post-lingua-franca world. TB argues that humanity is entering a stage where technology transcends Babel rather than reconstructing it: algorithms mediate understanding directly across languages, redefining what shared meaning means. However, the transformation remains paradoxical. While AI promises to democratize communication, it is still largely trained on Anglocentric data, potentially recoding English as an invisible algorithmic substrate rather than a spoken global standard. Drawing from sociolinguistics, media ecology, and AI ethics, this paper examines whether the digital age herald's linguistic liberation or the subtler persistence of English dominance within machine architectures. TB theory contends that the future of language will be determined not by which tongue prevails, but by how human and artificial intelligence chooses to interpret the diversity of expression.

Keywords: Trans-Babelism (TB), AI multilingualism, linguistic power, algorithmic bias, English dominance, epistemic justice, digital communication, cultural semiotics

iafor
The International Academic Forum
www.iafor.org

Introduction

Language is never just a system of signs and sounds, but a mechanism of power, a medium of identity, and a technology of governance. Human society have been accompanied by the rise of dominant languages: Latin in the Roman Empire, Arabic in the scholarly world of the medieval period, and more recently, English in the global era of industrialization, commerce and the internet. The ascendance of English is well documented as a process of "linguistic imperialism," the systematic privileging of one language over others through institutional structures, education systems and global networks (Alshahrani, 2020; Phillipson, 2018); this remains visible today in the persistence of English as the default lingua franca in science, business and diplomacy (Alshahrani, 2020; Lukianenko, 2024). In the 21st century, however, we are witnessing not just a continuation of English dominance (Zeng & Yang, 2024) but a potentially more radical transformation of the linguistic order driven by artificial intelligence (AI). With the proliferation of real-time translation tools, multilingual large language models (LLMs) and speech recognition systems, the possibility arises of a "post-lingua-franca" world in which no single human language must hold primacy (Ataman et al., 2025; Resck et al., 2025). AI technologies promise mediation across hundreds of languages, potentially dissolving the traditional hierarchy that placed English at the apex. AI-driven translation offers a nuanced understanding of the current state of language translation, emphasizing the advancements, challenges, and ethical considerations (Y. A. Mohamed et al., 2024). Yet this promise is complicated by structural realities. AI language systems are frequently trained on large English-centric corpora, evaluated with benchmarks skewed toward high-resource languages, and suffer from cultural and linguistic biases that favors Anglophone norms (Gao et al., 2024). While the interface of communication may appear to be democratizing, the underlying architecture may still embed English as the invisible substrate of meaning (Bender et al., 2021).

This paper proposes the Trans-Babelism (TB) theory to interpret this shifting landscape. The biblical myth of the Tower of Babel recounts humanity's attempt to unify under a single language and the subsequent divine scattering that produced linguistic plurality and mutual incomprehension (Genesis 11:1–9). Long interpreted as a symbolic origin of linguistic fragmentation, Babel has been invoked by scholars as a foundational metaphor for the entanglement of language, power, and meaning in human civilization (Eco, 1997; Steiner, 1975). In the era of AI, we might say that a new "tower" is emerging, not built of bricks but of algorithms, vectors and embeddings. The question is whether this new structure dismantles the old order, or merely re-constructs it in digital form under new terms (Crawford, 2022). TB argues that what will determine the future of language is not simply which tongue prevails, but how human and AI mediate diversity, whether the medium of understanding becomes multilingual equality or covert Anglophone hegemony.

Historical Context

English's rise was neither accidental nor purely linguistic. Its dominance was powered by colonial expansion, global trade, and later by digital communication. English became the world's lingua franca through institutional reinforcement rather than inherent superiority (Phillipson, 2018). The 20th century extended this influence into every domain, academia, diplomacy, business, and technology. Programming languages, scientific journals, and internet infrastructure all adopted English as their foundation. By the early 2000s, English was not merely a language but an infrastructure of globalization, a modern Babel (Crystal, 2003). Yet globalization also diversified English. The emergence of local variants both challenged and reinforced global linguistic hierarchies (Kachru, 1996). While English united people in

commerce and science, it also created new inequalities between native and non-native speakers, between linguistic centers and peripheries. Before English dominance, other languages held prominent roles. Latin served as the scholar's and cleric's lingua franca across medieval and early modern Europe: even when no longer a native tongue, Latin enabled communication among the educated across different vernaculars (Shank et al., 2013). French became lingua franca during the 17th to 19th centuries, replacing Latin in diplomacy and intellectual exchanges (Darquennes, 2016). Spanish achieved vehicle-language status through colonial expansion, embedding itself as a regional lingua franca even beyond Europe (Castillo, 2011). Russian played a similar intermediary-role in the 20th century as a language of communication across Central Asia and Eastern Europe (Pavlenko, 2006). Finally, while Mandarin Chinese serves as a national lingua franca within China, unifying numerous dialects and minority languages, it has so far not displaced English globally (Li, 2006). The emergence of AI promises to disrupt this pattern. Automated translation could decentralize linguistic authority, but the algorithms behind it remain shaped by the very dominance they appear to transcend.

Theoretical Foundations of Trans-Babelism

Sociolinguistics: Language as Power and Identity

Sociolinguistics conceives language not merely as a neutral conduit for communication but as a social institution that encodes hierarchy, power, and belonging. As Bourdieu (1991) observed, language operates as a form of symbolic capital, those who speak the "legitimate" or prestigious language gain access to social mobility, authority, and economic opportunity (Bourdieu, 1991). In this sense, linguistic competence is inseparable from social privilege. The global dominance of English thus extends far beyond grammar or vocabulary: it functions as a gateway to participation in international academia, diplomacy, and the digital economy (Canagarajah, 2012; Pennycook, 1994). Historically, the status of English emerged through colonial and post-colonial networks that linked education, religion, and governance. The spread of English in Africa, South Asia, and the Pacific institutionalized a linguistic hierarchy between "native" and "non-native" speakers, reinforcing what called linguistic imperialism (Phillipson, 2018). Even as World Englishes diversified, producing distinctive varieties in India, Nigeria, and Singapore, the symbolic hierarchy persisted: proximity to "Standard English" remained a marker of legitimacy (Kachru, 1996). In the age of AI, this hierarchy takes a new form. Language within AI systems becomes abstracted from human community and cultural experience, it becomes data. Every utterance, once a situated social act, is converted into quantifiable units: tokens, embeddings, and probabilities. Meaning is reconstructed through pattern recognition rather than lived context (Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, 2021). In doing so, linguistic authority shifts from the speaker to the system designer. The new "elite" are no longer those fluent in English but those who control the datasets, algorithms, and parameters that define what counts as coherent or correct language (Gao et al., 2024). This transition represents a profound reconfiguration of linguistic power. Where once discourse was negotiated through social interaction, it is now regulated through computational logic. The politics of accent, grammar, and idioms give way to the politics of training data, model architecture, and representational bias. In short, in the AI era, to own the corpus is to own the conversation. Sociolinguistic inequality persists, but its arena has shifted from classrooms and parliaments to servers and datasets.

Media Ecology: The Medium as Mediator

Media ecology examines how communication technologies shape not only what humans say but how they think, perceive, and organize society. The aphorism that "the medium is the message" encapsulates the foundational principle: every new medium reconfigures the scale and form of human association (McLuhan & Lapham, 1994). Language, in this framework, is both message and medium—a technology that extends the human mind. When spoken tongues evolved into writing, and writing into print, each transformation altered cognition, authority, and cultural continuity (Ong, 1982; Postman, 1970). The rise of AI represents the next phase in this trajectory: language itself becomes computational; its mediation automated through code. In digital modernity, communication no longer travels through purely human channels but through algorithmic infrastructures that filter, translate, and generate meaning. The user interacts not with another mind directly but through an intelligent interface that interprets and reformulates intention. This is like an entry into the Infosphere, where information objects—data, code, algorithms—constitute a shared environment of being (Floridi, 2014). Within this sphere, AI functions not as a neutral tool but as a new medium of mediation—a semi-autonomous agent that structures how meaning circulates. For example, machine translation tools such as Google Translate or DeepL do not simply carry messages from one language to another; they reshape the communicative act. They prioritize efficiency over ambiguity, legibility over cultural resonance. In effect, AI systems mediate the world through a "flattened semiotics" (Couldry & Mejias, 2020): they privilege what can be parsed by an algorithm, thereby standardizing linguistic diversity into a statistically tractable form. What McLuhan once observed of television, that it re-patterns attention and re-defines public discourse, now applies to AI language systems. They do not merely transmit human messages; they model how meaning itself should be structured.

The Trans-Babelism Framework

The first stage corresponds to the pre-modern multilingual world of Babel 1.0, characterized by local tongues, regional lingua francas, and reliance on human translators or interpreters. Before mechanical printing or global navigation, linguistic diversity was the norm rather than the obstacle. Communication across regions depended on cultural intermediaries—scribes, merchants, missionaries, whose work intertwined language with diplomacy and ritual (Burke, 2004). Latin served as the pan-European medium of theology and science, Arabic connected scholars across the Islamic world, and Sanskrit linked South and Southeast Asia through religion and law (Pollock, 2006). Yet these lingua francas remained bounded: they circulated within civilizational ecologies, not as universal systems. Translation was human, contextual, and slow, ensuring depth but limiting reach.

The second stage, Babel 2.0, emerges with industrial modernity and the expansion of European empires. English, propelled by colonialism and later by American cultural hegemony, became the de facto global lingua franca (Phillipson, 2018). The standardization of education, trade, and technology created a shared linguistic infrastructure that accelerated globalization but entrenched new inequalities (Pennycook, 1994). During this period, English evolved from a communicative tool into an epistemic filter: scientific publication, international law, and digital networks increasingly required proficiency in English for participation. As Kachru's "Three Circles" model illustrates, the Inner Circle (native English countries) controlled legitimacy, the Outer Circle (post-colonial states) adopted institutional English, and the Expanding Circle (the rest of the world) engaged English instrumentally (Kachru, 1996). This architecture of hierarchy mirrored global power relations. The "Babel 2.0" phase thus achieved what earlier

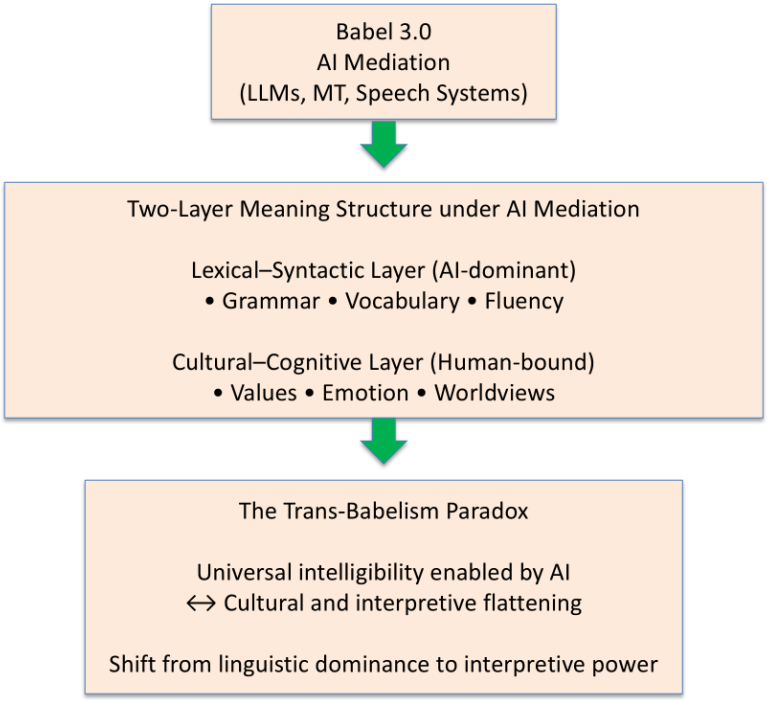
lingua francas never could but at the price of linguistic homogenization and epistemic exclusion. Languages without state or economic backing became marginalized, while English's symbolic capital (Bourdieu, 1991) deepened. In effect, humanity's dream of one shared tongue materialized through globalization, not divinity.

The third and current stage, Babel 3.0, introduces a qualitatively different paradigm: the mediation of language by artificial intelligence. Here, communication no longer depends on linguistic mastery or human translation but on computational models that convert speech, text, and images into mathematical representations. This stage appears to fulfill the ancient dream of transcending linguistic fragmentation. Real-time translation tools, large multilingual models, and speech-to-speech systems enable near-instant comprehension across hundreds of languages. In theory, every person can now speak in their own language and still be understood. Yet, as scholars warn, the infrastructure of this new Babel remains profoundly asymmetrical: most models are trained primarily on English data, Western cultural norms, and high-resource language pairs (Bender et al., 2021; Gao et al., 2024). Thus, Babel 3.0 represents not the end of English dominance but its metamorphosis. English becomes the hidden algorithmic substrate through which AI systems interpret and align other languages, a form of "computational lingua franca." The linguistic center has shifted from empire to algorithm, from classroom to cloud. The new priesthoods are not philologists but data scientists and engineers.

Figure 1 synthesizes the TB framework by illustrating how AI-mediated communication (Babel 3.0) restructures meaning into two asymmetric layers, giving rise to the TB Paradox, universal intelligibility alongside cultural and interpretive flattening.

Figure 1

Mechanism of Trans-Babelism Under AI Mediation



Two-Layer Model of AI-Mediated Communication

Layer 1: Lexical-Syntactic encompasses the structural and statistical dimensions of language: grammar, vocabulary, and surface semantics. Neural machine translation (NMT) systems and

large language models excel in this domain. They map tokens across vast multilingual corpora, identify probabilistic correspondences, and generate sentences that appear fluent and contextually appropriate (Vaswani et al., 2017).

The success of this layer is evident in real-time translation tools such as Google Translate, DeepL, and Meta's SeamlessM4T. These systems routinely achieve BLEU or COMET scores approaching human parity in high-resource language pairs like English–French or English–Spanish (Kocmi et al., 2023). Yet this computational fluency is contingent upon abundant parallel corpora, texts aligned sentence-by-sentence, and thus privileges dominant languages. The lexical-syntactic layer can be described as the quantitative frontier of communication: measurable, optimizable, and scalable.

However, this efficiency comes at a cost. Words are treated as interchangeable symbols rather than culturally situated acts. Semantic equivalence is defined statistically, not experientially. As noted, language models "simulate understanding without grounding"—they predict text forms but lack any connection to human intention or social meaning (Bender & Koller, 2020).

The second layer encompasses the pragmatic, emotional, and cultural contexts that imbue language with human depth. It includes metaphor, irony, politeness, and collective memory, the elements that resist quantification. In this realm, AI performance remains brittle. Research in cross-cultural pragmatics confirms this limitation. It has been found that neural models misinterpret indirect speech acts and politeness strategies because such cues are culture-specific, not syntactically encoded (Pavlick & Kwiatkowski, 2019). These shortcomings reveal that while AI can approximate meaning, it cannot inhabit it. In the cultural-cognitive layer, translation demands empathy, historical awareness, and shared worldviews, qualities that remain uniquely human. The distinction mirrors Jakobson's classic taxonomy of translation: from intralingual (within one language) to interlingual (between languages) and intersemiotic (between sign systems) (Jakobson, 1959). AI currently operates between the first two levels; the third, where symbols acquire affective resonance, remains inaccessible to computation.

These two layers are not isolated but interdependent. The lexical-syntactic layer provides efficiency; the cultural-cognitive layer provides authenticity. In human communication, both coexist seamlessly. In AI systems, however, they are dissociated—one automated, the other residual. Emerging research in neuro-symbolic and multimodal AI seeks to bridge this gap by combining linguistic data with visual and contextual cues (Bisk et al., 2020). Projects like Meta's No Language Left Behind (Costa-jussà, 2022) attempt to embed low-resource languages and cross-cultural representations. Yet these innovations remain technically driven rather than anthropologically grounded. Without explicit modeling of cultural cognition, the deeper layer of meaning continues to elude AI.

Case Illustrations: The Limits of Algorithmic Translation

Poetry is the ultimate stress test for any translation system because it compresses emotion, rhythm, and cultural resonance into condensed linguistic form. Machine translation systems tend to flatten these dimensions into semantic paraphrases. Scholars of translation studies emphasize that such poetic failure is not accidental but intrinsic: poetry depends on untranslatable, terms whose meanings are tied to sound, rhythm, or cultural metaphor (Apter, 2013). Neural systems cannot perceive aesthetic form or intertextual reference; they treat metaphor as statistical anomaly. Comparative analyses confirm this: while transformer-based models outperform earlier phrase-based systems in literal accuracy, human evaluators

consistently rate them poorly on figurative or emotional fidelity (Toral & Way, 2018). In this context, AI achieves what would call "interlingual translation" but fails at "intersemiotic translation," the movement between cognitive and affective universes (Jakobson, 1959). The TB framework interprets this failure as a sign of technological transcendence without human transcendence: a new tower that connects languages but not souls.

The translation of political rhetoric reveals another layer of AI's cultural inadequacy. Words like "freedom," "democracy," or "solidarity" carry distinct historical and moral weight in different societies. When AI translates such terms, it often normalizes them into Western liberal semantics, erasing ideological nuance. Translation of Arabic news headlines about gender or religion tends to reflect Western frames of secularism and individualism (Al-Mahrooqi & Denman, 2015). Such distortions illustrate the myth of algorithmic neutrality. AI systems trained on uneven global datasets reproduce the ideological balance of their sources. It has been noted that "every dataset carries a politics of representation"—the choice of what is translated, by whom, and for what purpose (Crawford, 2022). Thus, when AI claims impartial mediation, it actually reinforces an invisible Babel of epistemic asymmetry.

Humor epitomizes the cultural-cognitive layer because it relies on shared assumptions, double meanings, and timing. AI translation routinely fails to recognize irony, sarcasm, or wordplay, especially in cross-cultural contexts. Empirical studies confirm that humor is one of the hardest linguistic acts for AI to model. Even advanced transformer models achieve less than 50% accuracy in detecting sarcasm in multilingual corpora (Hossain et al., 2020). Likewise, computational humor systems can identify word associations but fail to grasp socio-cultural taboos or irony cues (Mihalcea & Strapparava, 2005). Humor depends on contextual intelligence, the tacit social knowledge of what is laughable, offensive, or absurd within a culture. AI, lacking experiential grounding, processes humor as lexical deviation rather than pragmatic performance. To translate perfectly, machines must first reduce meaning to what can be measured. In doing so, they risk converting the infinite variety of human expression into a "global village" (McLuhan & Lapham, 1994), connected yet homogenized. The challenge ahead is to design AI systems that re-embed culture within translation, drawing on cross-disciplinary collaborations between technologists, linguists, and anthropologists.

Implications for Global Communication

AI has begun to erode English's monopoly as a functional lingua franca while simultaneously entrenching it as an infrastructural code. English remains the dominant training language in machine translation, natural-language processing, and large-model corpora, constituting over 60 percent of web content and the majority of scientific repositories (W3Techs, 2025). Even when users communicate in Swahili, Vietnamese, or Arabic, the translation pathway often passes through an English pivot—effectively re-encoding global speech through an Anglophone filter (Tiedemann & Thottingal, 2020).

AI's ability to generate and translate across hundreds of languages fosters what may appear to be technological multilingualism—a surface pluralism in which all languages coexist digitally. UNESCO celebrates this potential as a pathway to inclusive education and cultural exchange (UNESCO, 2025). Yet scholars caution that automated translation often favors semantic clarity over cultural authenticity, producing what described as "commodified multilingualism": diversity repackaged for global consumption (Duchêne & Heller, 2012).

In education, AI translation tools democratize access to information. Students can now read textbooks in real time or converse with instructors across languages. However, research in bilingual cognition suggests that linguistic diversity itself fosters creativity, perspective-taking, and cognitive flexibility (Bialystok, 2018). If AI mediates all learning through an invisible English-based layer, learners may unknowingly internalize Anglophone logic even when reading in their native tongue. This has implications for global knowledge production. Non-English research, long marginalized in international citation systems (Hamel, 2007), risks further invisibility when AI prioritizes data from English-dominant sources. The resulting epistemic bias could perpetuate a "digital Matthew effect," where languages with large corpora grow ever more dominant while smaller linguistic traditions fade from scientific and intellectual discourse.

AI offers powerful tools for revitalizing endangered languages, speech synthesis for Māori, corpus building for Yoruba, or digital archives for Ainu (Bird, 2020). Yet such projects raise issues of linguistic sovereignty: Who owns the model trained on indigenous data? Who profits from its deployment? The case of the Māori-language chatbot Kōrerorero (2022) exemplifies both promise and peril, community engagement preserved nuance, but commercialization risked appropriation.

The final implication of TB is philosophical. If AI becomes the universal interpreter, then ethics must address not only who speaks but how machines listen. The goal is not the erasure of difference but its intelligible coexistence: an AI pluralism where translation preserves worldview as much as word. As reminded to translators decades ago, "translation is not words for words but meanings for people." In the age of intelligent machines, that humanistic reminder becomes newly urgent (Nida, 2002).

Conclusion

The biblical story of Babel symbolizes humanity's tension between unity and diversity. The rise of AI is the most ambitious attempt yet: transcending Babel not by enforcing one tongue, but by allowing all languages to speak through one machine. Yet TB theory suggests this achievement is both revolutionary and illusory. AI removes linguistic barriers but replaces them with algorithmic ones. The hierarchy of languages becomes a hierarchy of models, datasets, and architectures. English, once the language of global commerce, now reappears as the hidden code of computation. Communication becomes instant, yet filtered through systems few understand. The miracle of translation risks becoming the mirage of comprehension. Still, this paradox offers possibility. If Babel 1.0 fragmented humanity and Babel 2.0 unified it under English, Babel 3.0 may enable coexistence through interpretation. True transcendence will not come from perfect translation but from ethical mediation, technologies that preserve difference as a source of meaning. This requires moving from linguistic justice to interpretive justice. AI must be accountable not only for translation but for understanding, while policymakers, educators, and technologists ensure that all languages retain their epistemic dignity. Ultimately, the future of language depends on how humanity chooses to listen. The goal is not to rebuild Babel in code but to transcend it in spirit, recognizing diversity as humanity's collective intelligence. In this vision, TB marks not an end, but a beginning.

References

Al-Mahrooqi, R., & Denman, C. (2015). *Issues in English education in the Arab world*. Cambridge Scholars Publishing.

Alshahrani, A. (2020). Linguistic Imperialism and the English Language: A Review of Literature. *Ad Alta Journal of Interdisciplinary Research*, *10*(1), 8–10. https://www.researchgate.net/publication/343166363

Apter, E. (2013). *Against World Literature: On the Politics of Untranslatability*. Verso.

Ataman, D., Birch, A., Habash, N., Federico, M., Koehn, P., & Cho, K. (2025). Machine Translation in the Era of Large Language Models: A Survey of Historical and Emerging Problems. *Information*, *16*(9). https://doi.org/10.3390/info16090723

Beer, D. (2019). *The Data Gaze: Capitalism, Power and Perception*. SAGE Publications Ltd. https://doi.org/10.4135/9781526463210

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bender, E. M., & Koller, A. (2020). Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 5185–98.

Bialystok, E. (2018). Bilingual education for young children: review of the effects and consequences. *International Journal of Bilingual Education and Bilingualism*, *21*(6), 666–679. https://doi.org/10.1080/13670050.2016.1203859

Bird, S. (2020). Decolonising Speech and Language Technology. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 3504–3519). International Committee on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.313

Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., Lapata, M., Lazaridou, A., May, J., Nisnevich, A., Pinto, N., & Turian, J. (2020). Experience Grounds Language. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 8718–8735). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.703

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in {NLP}. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485

Bourdieu, P. (1991). *Language and Symbolic Power*. Harvard University Press.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability, and Transparency*, 77–91.

Burke, P. (2004). *Languages and Communities in Early Modern Europe*. Cambridge University Press.

Canagarajah, S. (2012). *Translingual Practice: Global Englishes and Cosmopolitan Relations* (1st ed.). Routledge.

Castillo, M. (2011). Globish and the empire. *American Journal of Neuroradiology, 32*(8), 1363–1364. https://doi.org/10.3174/ajnr.A2303

Costa-jussà, M. (2022). No Language Left Behind: Scaling Human-Centered Machine Translation. *ArXiv Preprint*, 2207.04672.

Couldry, N., & Mejias, U. A. (2020). The Costs of Connection: How Data Are Colonizing Human Life and Appropriating It for Capitalism. *Social Forces*, *99*(1), e6–e6. https://doi.org/10.1093/sf/soz172

Crawford, K. (2022). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.

Darquennes, J. (2016). 2.3 Lingua francas of Europe. In *Contexts and Agendas* (pp. 28–33). De Gruyter Mouton. https://doi.org/doi:10.1515/9781614518952-006

Duchêne, A., & Heller, M. (2012). *Language in Late Capitalism: Pride and Profit*. Routledge.

Eco, U. (1997). *The Search for the Perfect Language*. Fontana Press.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Gadamer, H. (1975). Truth and Method. *Journal of Aesthetics and Art Criticism*, *36*(4), 487–490.

Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., & Li, Y. (2024). Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Humanities and Social Sciences Communications*, *11*(1), 1259. https://doi.org/10.1057/s41599-024-03611-3

Hamel, R. E. (2007). The dominance of English in the international scientific periodical literature and the future of language use in science. *AILA Review*, *20*(1), 53–71.

Hossain, N., Krumm, J., Gamon, M., & Kautz, H. (2020). {S}em{E}val-2020 Task 7: Assessing Humor in Edited News Headlines. In A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, & E. Shutova (Eds.), *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 746–758). International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.98

Jakobson, R. (1959). On Linguistic Aspects of Translation. In B. R. (Ed.), *On Translation* (pp. 232–239). Harvard University Press.

Kachru, B. B. (1996). World Englishes: Agony and Ecstasy. *Journal of Aesthetic Education*, *30*(2), 135–155. http://www.jstor.org/stable/3333196

Kocmi, T., Avramidis, E., Bawden, R., Bojar, O., Dvorkovich, A., Federmann, C., Fishel, M., Freitag, M., Gowda, T., Grundkiewicz, R., Haddow, B., Koehn, P., Marie, B., Monz, C., Morishita, M., Murray, K., Nagata, M., Nakazawa, T., Popel, M., … Suzuki, J. (2023). Findings of the 2023 Conference on Machine Translation ({WMT}23): {LLM}s Are Here but Not Quite There Yet. In P. Koehn, B. Haddow, T. Kocmi, & C. Monz (Eds.), *Proceedings of the Eighth Conference on Machine Translation* (pp. 1–42). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.1

Li, D. C. S. (2006). Chinese as a lingua franca in Greater China. *Annual Review of Applied Linguistics*, *26*, 149–176. https://doi.org/10.1017/S0267190506000080

Lukianenko, N. (2024). Language and power: linguistic imperialism. *International Science Journal of Education & Linguistics*, *3*(5), 41–49. https://doi.org/10.46299/j.isjel.20240305.06

McLuhan, M., & Lapham, L. H. (1994). *Understanding Media: The Extensions of Man*. The MIT Press.

Mhlambi, S., & Tiribelli, S. (2023). Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms. *Topoi*, *42*(3), 867–880. https://doi.org/10.1007/s11245-022-09874-2

Mihalcea, R., & Strapparava, C. (2005). Computational Laughing: Automatic Recognition of Humorous One-liners. *Current Opinion in Behavioral Sciences*, *29*, 89–94.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H. M., Adiel, M. A. E., & Elsadig, M. A. (2024). The Impact of Artificial Intelligence on Language Translation: A Review. *IEEE Access*, *12*, 25553–25579. https://doi.org/10.1109/ACCESS.2024.3366802

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Fagbohungbe, T., Akinola, S. O., Muhammad, S., Kabongo Kabenamualu, S., Osei, S., Sackey, F., Niyongabo, R. A., Macharm, R., Ogayo, P., Ahia, O., Berhe, M. M., Adeyemi, M., Mokgesi-Selinga, M., Okegbemi, L., Martinus, L., … Bashir, A. (2020). Participatory Research for Low-resourced Machine Translation: A Case Study in {A}frican Languages. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2144–2160). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.195

Nida, E. A. (2002). *Language and Culture: Contexts in Translating.* Shanghai Foreign Language Education Press.

Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge.

Pavlenko, A. (2006). Russian as a lingua franca. *Annual Review of Applied Linguistics*, *26*, 78–99. https://doi.org/10.1017/S0267190506000055

Pavlick, E., & Kwiatkowski, T. (2019). Inherent Disagreements in Human Textual Inferences. *Transactions of the Association for Computational Linguistics*, *7*, 677–694. https://doi.org/10.1162/tacl_a_00293

Peirce, C. S. (1931). *Collected Papers of Charles Sanders Peirce*. Harvard University Press.

Pennycook, A. (1994). *The cultural politics of English as an international language* (1st ed.). Routledge. https://bayanebartar.org/file-dl/library/Linguistic1/The_Cultural_Politics_of_English.pdf

Phillipson, R. (2018). Linguistic Imperialism. In *The Encyclopedia of Applied Linguistics* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781405198431.wbeal0718.pub2

Pollock, S. (2006). *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India* (1st ed.). University of California Press. http://www.jstor.org/stable/10.1525/j.ctt1pnqs7

Postman, N. (1970). The reformed English curriculum. In *High school 1980: The shape of the future in American secondary education*. Pitman.

Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the Everything in the Whole Wide World Benchmark. *Advances in Neural Information Processing Systems*, *NeurIPS*.

Resck, L., Augenstein, I., & Korhonen, A. (2025). *Explainability and Interpretability of Multilingual Large Language Models: A Survey*. *Section 4*, 20465–20497. https://doi.org/10.18653/v1/2025.emnlp-main.1033

Shank, M. H., Lindberg, D. C., & Shank, M. H. (2013). Schools and Universities in Medieval Latin Science. In *The Cambridge History of Science* (pp. 207–239). Cambridge University Press.

Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating Gender Bias in Machine Translation. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1679–1684). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1164

Steiner, G. (1975). *After Babel: Aspects of Language and Translation*. Oxford University Press.

Tiedemann, J., & Thottingal, S. (2020). {OPUS}-{MT} {--} Building open translation services for the World. In A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg, & M. L. Forcada (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 479–480). European Association for Machine Translation. https://aclanthology.org/2020.eamt-1.61/

Toral, A., & Way, A. (2018). *What Level of Quality Can Neural Machine Translation Attain on Literary Text?* 263–287. https://doi.org/10.1007/978-3-319-91241-7_12

UNESCO. (2025). *Languages matter: global guidance on multilingual education*. https://www.unesco.org/en/articles/languages-matter-global-guidance-multilingual-education

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

W3Techs. (2025). *Usage statistics of content languages for websites*. https://w3techs.com/technologies/overview/content_language

Zeng, J., & Yang, J. (2024). English language hegemony: retrospect and prospect. *Humanities and Social Sciences Communications*, *11*(1), 317. https://doi.org/10.1057/s41599-024-02821-z

Contact email: tran@tokushima-u.ac.jp