3-25-2024

Preserving Linguistic Diversity in the Digital Age: A Scalable Model for Cultural Heritage Continuity

James Hutson
Lindenwood University, jhutson@lindenwood.edu

Pace Ellsworth
Preserving Records Everywhere (PRE)

Matt Ellsworth
Preserving Records Everywhere (PRE)

Article

Preserving Linguistic Diversity in the Digital Age: A Scalable Model for Cultural Heritage Continuity

James Hutson[1,*] iD , Pace Ellsworth[2] iD , and Matt Ellsworth[3] iD

[1]Department Head of Art History and Visual Culture, College of Arts and Humanities, Lindenwood University, Saint Charles, MO, USA
[2]CEO, Preserving Records Everywhere (PRE), BA Computational Linguistics, Woodridge, IL USA
[3]VP of Content, Preserving Records Everywhere (PRE), MA Spanish Translation, Pattonsburg, MO USA

* Corresponding author: James Hutson, Department Head of Art History and Visual Culture, College of Arts and Humanities, Lindenwood University, Saint Charles, MO, USA. Email: jhutson@lindenwood.edu

ARTICLE INFO

ABSTRACT

In the face of the rapid erosion of both tangible and intangible cultural heritage globally, the urgency for effective, wide-ranging preservation methods has never been greater. Traditional approaches in cultural preservation often focus narrowly on specific niches, overlooking the broader cultural tapestry, particularly the preservation of everyday cultural elements. This article addresses this critical gap by advocating for a comprehensive, scalable model for cultural preservation that leverages machine learning and big data analytics. This model aims to document and archive a diverse range of cultural artifacts, encompassing both extraordinary and mundane aspects of heritage. A central issue highlighted in the article is the dominance of English, especially in popular culture, and its role in reshaping cognitive patterns, leading to the loss of traditional languages and cultures. This trend, favoring orality over literacy, amplifies the need for the preservation actions proposed. The article underscores the importance of establishing a more meaningful and comprehensive lingua franca as the first step in this endeavor. This approach is not just about safeguarding languages but also about preserving the nuanced ways in which different linguistic paradigms shape our understanding of the world. The paper explores how language evolution, translation nuances, and the unique perspectives inherent in various linguistic paradigms contribute to a rich cultural diversity. It also examines the impact of technology and social media on traditional cultural norms and the critical need for deliberate efforts to reinforce endangered languages and practices. The proposed model aims to bridge the gap between past and present cultural expressions, using technology to reinforce mother-tongue contexts and enhance the appreciation of diverse human identities and worldviews. Through this, the article presents a compelling case for a more inclusive and holistic approach to preserving the global tapestry of cultural heritage.

1. Introduction

The linguistic diversity of our world is both immense and fragile. Currently, there are approximately 7,151 languages spoken globally, yet this diversity faces significant threats. Since 1900, about 10% of these languages, roughly 700, have been lost (Park, 2023). Alarmingly, since 2000, languages have been disappearing at a rate of one every 40 days, equating to nine languages annually (Aalberse et al., 2019). Projections indicate a dire situation: around 3,000 languages could disappear by the end of this century, at a rate of one every two weeks. This loss signifies not just the silencing of voices but the erasure of rich cultural heritages and unique world perspectives (Kandler & Unger, 2023). As modernity advances, the variety and distinction of our languages, and the cultures they represent, are diminishing rapidly. After global warming, The Language Conservancy argues, language loss is emerging as Earth's most acute crisis (Collette & Kennedy, 2023).

The dominance of English in global communication and

its implications for linguistic diversity is a pivotal issue that requires careful examination. Despite the frequent mention of "language" in discussions on lingua franca, the article often falls short of specifically addressing the overarching influence of English. This lack of specificity overlooks the nuanced effects of the predominance of the language on cognitive patterns and cultural assimilation processes. For instance, Crystal (2003) highlights the role of English as a global lingua franca, impacting linguistic diversity and leading to the marginalization of minority languages. Similarly, Phillipson (1992) critiques the linguistic imperialism of English, suggesting that its dominance perpetuates inequalities and hinders the preservation of linguistic heritage. Additionally, Skutnabb-Kangas (2013) emphasizes the risks to linguistic human rights posed by the dominance of major languages like English. Graddol (1997) forecasts the future dynamics of English as a global language, predicting its continued expansion and the challenges this poses for linguistic diversity. The integration of these perspectives underscores the complex interplay between the global dominance of the Eurocentric language and the urgent need for strategies to safeguard linguistic diversity, illustrating the critical gap in the article's discussion.

Recent research on language loss has highlighted the accelerating rate at which languages are disappearing and the potential consequences of this loss (Shcherbakova et al., 2023). Skirgård et al. (2023), for instance, used the largest database of language structure in the world, Grambank, to understand global linguistic diversity and the threat posed by language loss. The study calculated the potential effects of language loss globally and regionally, using a metric called "functional richness" to measure the consequences of language loss in different regions. Another study by Sefotho et al. (2023) reported that grammatical structure is highly flexible across languages and that language extinction has been accelerating due to social, political, and economic pressures. The study emphasized the importance of sustained efforts to document and revitalize endangered languages to preserve the linguistic window of humanity into history, cognition, and culture. Furthermore, Bromham et al. (2022) warned that without intervention, language loss could triple within 40 years, with at least one language being lost per month. The study highlighted the urgent need for investment in language documentation, bilingual education programs, and other community-based initiatives to avoid the loss of over 1,500 languages by the end of the century.

These studies underscore the criticality of documenting and revitalizing endangered languages, crucial in safeguarding human history and culture. Language diversity is essential in preserving cultural heritage, acting as a medium for conveying culture, values, beliefs, and history across generations (Sevinç, 2022). Each language is a unique tool for expressing identities and shared experiences. The words, phrases, and expressions in different languages hold the accumulated wisdom and experiences of their speakers (Jerome et al., 2022). The loss of a language means the loss of this invaluable knowledge. Hence, preserving rare languages is not just vital for linguistic studies but also for maintaining cultural depth. Central to a community's cultural identity, language ensures the continuation of traditions, stories, and values. Documenting, preserving, and revitalizing endangered languages are essential steps to protect linguistic diversity and the cultural heritage they represent.

The predictions about the disappearance of languages could be further impacted by the fastest technological adoption in history. The rise of Big Tech, especially in the United States, has fostered a global shift towards English as the dominant lingua franca. This trend is exacerbated by the development of generative AI by major tech companies like Microsoft, Google, Meta, Tesla, and Amazon. These companies, largely based in the U.S., are at the forefront of creating large language models (LLMs) that predominantly use English (Lai et al., 2023). This focus on English in the development of LLMs poses significant risks. While these technologies have the potential to enhance communication and information access globally, their English-centric design could inadvertently contribute to the erosion of linguistic diversity (Kirk et al., 2023; Zhang et al., 2023). By prioritizing English, these models may not only overlook but also devalue other languages, especially those that are less commonly spoken or at risk of extinction (Chang et al., 2023; Salomone & Salomone, 2022). According to Ramesh et al. (2023), this situation could lead to a further decline in the use of these languages, as people increasingly adopt English to access and leverage technology effectively.

Additionally, a trend towards orality over literacy is unfolding. As natural language communication with computers evolves, particularly through large language models, a significant shift from traditional literacy towards orality is emerging. This transition underscores the adaptability and immediacy of spoken interactions with technology, mirroring human-to-human communication patterns. Bers (2022) has addressed this phenomenon, examining how programming education for children can transition from traditional written communication to more oral and interactive methods, reflecting broader societal shifts in communication preferences. Their work highlights the natural inclination of young learners towards oral communication, suggesting a pedagogical shift that could influence future language learning and technology interaction strategies. This evolution towards orality, facilitated by advances in natural language processing, presents both challenges and opportunities in maintaining the richness of linguistic diversity in digital spaces. The integration of these insights from recent literature highlights the ongoing dialogue between technology and language, illustrating the dynamic nature of communication in the digital age. In essence, the technological advancements spearheaded by these major companies, while beneficial in many ways, may inadvertently contribute to the Americanization of global communication and culture (Konieczny & Lewoniewski, 2023). This trend highlights the need for a more inclusive approach in the development of AI and technology, one that acknowledges and supports the preservation of linguistic and cultural diversity worldwide.

As such, this paper explores linguistic diversity and cultural heritage preservation, highlighting how language, in the broadest sense of the term, is central to these aspects. It will address the emergence of a new Lingua Franca through globalization and linguistic homogenization, emphasizing the dominance of English and its implications for cultural diversity. The goal is to spur action among scholars, linguists, governments, cultural heritage organizations, and communities. By acknowledging the perceived supremacy of Occidental culture and the consequent loss of other languages, the paper aims to provide AI and computational linguistics-based recommendations to counteract these trends and promote the preservation of diverse linguistic heritages.

2. Theoretical Background

The theoretical background for the study of linguistic diversity and technology integration is deeply rooted in the understanding of how digital advancements influence language use and preservation. Cruz (2024) explores the integration of artificial intelligence tools in education, emphasizing the importance of considering cultural and linguistic diversity. This approach aligns with the work of Torres-Zúñiga and Söğüt (2024), who investigate the use of intercultural virtual exchange to enhance English language teachers' critical pedagogy practices, highlighting the potential of technology to support diverse linguistic experiences in the classroom. Additionally, Morgenstern and Boutet (2024) contribute to this theoretical framework by examining the orchestration of bodies and artifacts in social settings, suggesting how digital technologies might integrate into everyday practices to support linguistic diversity. These recent studies underscore the complex interplay between technology and language, suggesting that technological advancements can both challenge and enrich the tapestry of global linguistic diversity.

From a biological and neuroscience perspective, bilingualism and multilingualism are indicators of robust cognitive, mental, and physical health (Koch et al., 2023). Research shows that individuals who speak multiple languages exhibit greater cognitive flexibility, improved problem-solving skills, and enhanced memory function (Andleeb et al., 2023). Neuroscientific studies reveal that bilingualism can delay the onset of dementia and contribute to the neuroplasticity of the brain (Gallo & Abutalebi, 2023). Kazanas and Altarriba (2012) highlighted the buffering effect of multilingualism on memory loss in the elderly and its potential to improve executive functioning. Additionally, Diamond (2010) emphasized that bilingual children develop specific cognitive benefits during infancy, offering some protection against symptoms of Alzheimer's dementia in older people. Furthermore, the process of learning and using multiple languages can lead to increased mental agility and resistance to cognitive decline with age.

Reviewing literature on language loss and the disappearance of bilingualism and multilingualism reveals a multifaceted and complex phenomenon influenced by various factors, including mental health, brain injury, social contexts, and cognitive development (Alexandrova, 2023; Ng & Wigglesworth, 2007; Szubko-Sitarek, 2015). One aspect of language loss in bilingual individuals is its association with mental health. El-Gabalawi and Khouzam (2009) found that affective disorders and acute exacerbation of preexisting mental illnesses can lead to language regression, with the second language being more vulnerable to the effects of mental illness than the native language. At the same time, brain injuries can significantly impact language abilities in bilingual and multilingual contexts. Studies have shown that language impairment in bilingual speakers with post-stroke aphasia is influenced by factors such as age of language acquisition, frequency of language use, and premorbid proficiency (Kuzmina et al., 2019). Obler et al. (2012) noted that various patterns of language loss are linked to the damaged brain regions in multilingual individuals.

In addition to cognitive benefits, there are also indications of social and cultural advantages, enriching personal and community experiences. The social and linguistic environment plays a significant role in language maintenance or loss (Hollebeke et al., 2023). On the other hand, language loss can broadly be identified not as the result of a biological mechanism, but rather cultural ones. Research reveals a complex interplay of factors influencing language maintenance, shift, and loss. Research highlights the multifaceted dynamics at play in the preservation, shift, and disappearance of languages. Blackledge et al. (2015) underline the critical role of cultural context in determining whether a language is maintained or lost in environments where multiple languages are spoken. This insight is supported by Anderson (1999), who discovered that the loss of language among bilingual children is more likely to result from the contexts in which different languages come into contact with each other, rather than from any deficiencies in the children's ability to learn languages. Doiz et al. (2014) discuss the impact of globalization and internationalization in higher education on language choices, potentially leading to language loss. Lanza and Svendsen (2007) illustrate how language choices in bilingual and multilingual individuals are influenced by their social networks and identity. Hale (2015) highlights the loss of cultural and intellectual diversity due to the dominance of politically dominant languages over indigenous languages. Grohmann et al. (2016) explore how bilingualism impacts language acquisition and development within sociolinguistic frameworks. Martinet al. (2019) underscore the importance of cultural and linguistic compatibility in caregiving for maintaining language and well-being. These dynamics of language loss in bilingual and multilingual contexts are deeply intertwined with social and cultural factors, including the influence of cultural context, socio-cultural struggles, identity, political dominance, and sociolinguistic factors. Understanding these factors is crucial for addressing language loss and supporting linguistic diversity.

3. Challenges and Concerns

The challenges associated with preserving linguistic

diversity are multifaceted and encompass a range of issues from technological limitations to sociocultural barriers. Wang (2024) discusses the role of artificial intelligence in bilingual education, identifying the potential challenges in applying AI technologies to language preservation, including ensuring textual integrity and addressing the complexities of diverse linguistic contexts. Zhang, Jiang, Liu, and Chen (2024) propose data augmentation as a potential solution to the scarcity of linguistic data, emphasizing the importance of maintaining accuracy while enhancing the diversity of language samples. Furthermore, Dildora and Tukhtakhodjayeva (2024) highlight the socio-cultural challenges in preserving formulaic language expressions across different languages, stressing the importance of maintaining the socio-cultural essence during the translation process. These studies collectively underscore the critical need for innovative solutions that address both the technological and cultural dimensions of linguistic diversity preservation.

In addressing the challenge of language loss, recent scholarly work has identified several effective strategies to preserve linguistic diversity in bilingual and multilingual contexts. Wei (2014) underscores the importance of nurturing multilingual practices and social spaces where multiple languages are valued and actively used. Similarly, Blackledge et al. (2018) advocate for culturally inclusive pedagogy, emphasizing the need for teaching methods that respect and incorporate the linguistic backgrounds of students. Macswan (2017) presents an integrated multilingual model, viewing bilingualism as a norm and promoting the integration of multiple languages in various domains. In the professional realm, Van der Worp et al. (2017) propose recognizing and valuing linguistic diversity in workplaces, while Centeno (2009) suggests a multidisciplinary approach in healthcare to support language maintenance. Finally, Benson (2010) focuses on the importance of developing flexible and relevant multilingual strategies in education, tailoring teaching and learning practices to the multilingual realities of learners. These diverse strategies collectively contribute to the preservation of linguistic diversity, underscoring the importance of creating environments where multiple languages can thrive.

The limitations of previous efforts to counteract language loss need be emphasized. These strategies, though effective to some extent, often encounter difficulties in practical application and broad-scale implementation, particularly in the context of rapid global changes. Moreover, these approaches have not completely addressed the intricate interaction of technological advancement, globalization, and socio-political factors affecting linguistic diversity. In light of these challenges, the subsequent sections of this paper will delve into innovative approaches and fresh perspectives, aiming to offer a more holistic solution for preserving linguistic diversity and cultural heritage amidst modern complexities.Before discussing strategies to preserve languages that are disappearing, it is necessary to underscore the linguistic limitations of the English language, which, despite its widespread use and rich

vocabulary, may not be the ideal candidate for a universal language. English, with its origins rooted in a blend of Germanic and Romance languages, presents a unique linguistic profile. This amalgamation has endowed English with a rich and nuanced vocabulary, but it also contributes to its complexity and irregularities. For instance, English grammar and spelling are often seen as inconsistent due to their diverse origins. Words borrowed from different languages retain their original spelling and pronunciation rules, leading to a language that is less phonetically consistent compared to others (Baugh & Cable, 1993).

Moreover, the structure and syntax of English can be challenging for learners, especially those whose native languages have different grammatical systems. The idiosyncrasies in English grammar, such as irregular verb forms and the use of articles, can be particularly perplexing (Hinkel, 2001). This complexity does not inherently disqualify English as a global language, but it highlights the challenges non-native speakers face in acquiring proficiency. In terms of communication strategies and meaning-making, the strengths of the language lie in its expressive capacity and adaptability. It has a vast array of idiomatic expressions and a capacity for precise expression in various fields, from science to literature (Myers-Scotton, 2003). However, this richness can also be a barrier, as the subtleties of meaning and context-dependent nuances might not translate easily across cultures.

Furthermore, the evolution of languages like Old Norse and Middle English into their modern forms shows the dynamic nature of language. This evolution, while a natural linguistic process, can create distance between the original intent and meaning of words and their modern usage (Kroch et al., 2000). In English, this is compounded by its borrowing from multiple sources, making it a complex linguistic tapestry that might not always adhere to consistent rules. Thus, while the richness and global reach of the language are undisputed, its complex origins, irregular grammar, and nuanced vocabulary present challenges. These factors, combined with the natural evolution of language, suggest that English, like any language, has limitations in its role as a universal linguistic medium (Makhmudov, 2023).

These considerations underscore the limitations of English, but so too does the significance of understanding the context of words and phrases in a language, which parallels the importance of contextualizing artifacts in cultural heritage. Just as museum exhibits risk losing meaning when displayed out of context, languages lose their richness and depth when removed from their cultural and historical settings (Chen et al., 2023). Words and phrases like *Schadenfreude* or *ennui*, adopted into English, demonstrate the limitations of a single language in capturing the full spectrum of human experience. These loanwords reflect the necessity of embracing linguistic diversity to enhance mutual understanding and enrich all languages, and, at the same time, note the limitations of English to capture meaning (Park)nson, 2023).

At the same time, the adoption of English as a universal language has considerations that move beyond the linguistic. The shift also has profound psychological effects,

particularly in terms of alienation from one's mother tongue. This alienation is not just a loss of language but a disconnection from cultural roots and traditional knowledge (Savaya et al., 2023). For speakers of languages like Cherokee or Navajo, the inability to fully express themselves in their native tongue leads to a loss of cultural identity and a disconnect from ancestral wisdom (Gannie, 2023). This phenomenon is exacerbated by technology, which often prioritizes English and further marginalizes other languages. The result is a generation growing up linguistically and culturally disconnected, constantly seeking semantic context in a rapidly evolving digital landscape (Ochieng & Waithanji Ngware, 2023). Similarly, the way museums often present artifacts out of context, stripping them of their cultural and historical significance, is analogous to the decontextualization of languages. Just as Victorian-era aesthetics in museums can exoticize and commodify non-Western cultures, the dominance of English can similarly reduce the richness of other languages to mere novelties or linguistic commodities (Ahlqvist & Potter, 2023). This parallel emphasizes the need for a broader understanding and appreciation of linguistic and cultural contexts, recognizing their intrinsic value beyond mere curiosities or tools for communication.

Now that the nuances of linguistic meaning have been discussed, the focus shifts to the challenges of capturing the essence of languages through computational methods. For instance, the core of language preservation lies not just in recording what is spoken, but in understanding the original intent and context when a language was used. This involves delving into the historical and cultural underpinnings of a language to grasp what the speakers were truly conveying (Tankosić & Dovchin, 2023). For example, the concept of 'understanding' differs across languages – in English, it is to 'stand under', while in Russian, it implies 'taking up' or 'comprehending', and in German, it translates to *Verstehen*. Each translation carries a deep cultural and cognitive difference, highlighting the importance of understanding the mindset behind the language.

4. Discussion of Model

In order to trace these various branching meanings in languages, researchers often turn to computational linguistics. The use of semantic vector databases in linguistics, involving millions of dimensions of meaning, offers insights into how different languages relate to each other. These databases can predict relational meaning maps that are surprisingly similar across languages and even between humans and animals (Rataj et al., 2023). For instance, by altering relational terms, the concept of a 'dog' can transform into a 'cat', demonstrating the fluidity and interconnectedness of language concepts. However, preserving the uniqueness of each language requires an understanding of these complex relational maps. While AI and computational linguistics have made significant strides, they still struggle to capture the deeper meanings of words and phrases. Corpus-based translations, dictionary translations, and even recent generative translations often

miss the nuanced intent behind words (Zhang & Zhu, 2023). This limitation is particularly evident in languages where the connection between sound and meaning differs from that in English. AI often relies on a lone-word language model, where each word is memorized without much grammatical sense-making, leading to gaps in conveying the true essence of non-English languages (Sharma et al., 2021).

The goal of language preservation is not to halt innovation or evolution but to ensure that as languages evolve, there is a conscious effort to access and understand the wealth of knowledge and perspective from the past. This requires an anthropological approach, where the intelligence and context of past language users are considered. By integrating this historical perspective with modern technological advancements, it is possible to preserve languages in a way that honors their original intent and cultural significance (Govia, 2020). Thus, the preservation of languages involves a deep understanding of the original context and intent behind words and phrases. Computational methods like vector databases offer promising tools, but they must be used in conjunction with a nuanced understanding of linguistic and cultural history. The challenge lies in balancing the need for innovation with the importance of maintaining a connection to the linguistic and cultural heritage of past generations.

To effectively address the challenges of preserving linguistic diversity and cultural heritage, a combination of immediate and forward-looking strategies is essential. The first step involves establishing a more meaningful and comprehensive lingua franca. This new form of universal language would function as a general API between all languages, facilitating deeper understanding and connection across different linguistic communities and should be designed to capture the nuances and cultural contexts of various languages, thereby enhancing global communication. Simultaneously, there is an immediate need to use existing tools and technologies to reinforce cultural and linguistic traditions. This strategy focuses on preserving the context and richness of mother tongues, employing AI and machine learning not just for translation, but for supporting and revitalizing languages at risk of disappearing. The goal is to ensure these languages continue to be actively used, appreciated, and integrated within their native cultural contexts.

Moreover, a scalable, data-driven model for cultural preservation can be developed, adaptable to various cultural contexts and scopes. This model would incorporate machine learning and big data analytics to enable a nuanced capture and archiving of both extraordinary and everyday cultural artifacts (Basu et al., 2023). The first practical application could involve processing over a million records from specific cultural contexts, such as Burundi, serving as a large-scale case study (Hutson et al., 2023). This approach not only preserves linguistic diversity but also ensures the continued vibrancy and relevance of cultural heritage in an increasingly digital and interconnected world. However, there remains a dominant narrative in AI research on how distinct LLM are from how humans think, which is inhibiting the creation of such a model.

The proposed model, the Holistic Archival Personality Profiling Model (HAPPM), represents a groundbreaking approach in the field of cultural heritage preservation, particularly in the context of Bujumbura, Burundi (Hutson et al., 2023).. This innovative model addresses the urgent need to preserve and revitalize cultural identities through the digitization and semantic classification of a broad spectrum of cultural artifacts. By leveraging advanced digital methodologies, including machine learning (ML), natural language processing (NLP), and large language models (LLMs), HAPPM facilitates the construction of a comprehensive "space-time continuum" for individuals and communities. This multi-disciplinary approach enables the archiving of not only government records and notable events but also personal correspondences and family memorabilia, providing a richer, more nuanced view of cultural identities.

HAPPM's strength lies in its ability to bring together diverse data forms, transforming them into a digitally accessible format that is organized using biotags, chronotags, and geotags. This not only preserves the cultural heritage but also allows for the recreation of spoken languages and local dialects, enriching the appreciation of diverse human identities. The model's application in Bujumbura showcases its potential to bridge the gap between past and present cultural expressions, reinforcing native languages and contributing to a more inclusive method for preserving the global tapestry of cultural heritage. Through the HAPPM approach, the study demonstrates how technology can be harnessed to safeguard and revive cultural heritage, offering new avenues for understanding and appreciating the complexity of human cultures in an increasingly digital world.

The model also allos for further investigation into the notion of "Absent Presence," which becomes a pivotal concept for understanding how AI mirrors human linguistic capabilities and creativity. This term, drawing from Derridian deconstructive theory and the broader landscape of technoculture, encapsulates how AI, though seemingly autonomous, is deeply rooted in human influence and contribution (Lucia, 2023). At the core of this discussion is the acknowledgment that AI, particularly LLMs like ChatGPT, function as an extension of human cognition and creativity. These models, trained on vast arrays of human-generated content, encapsulate a broad spectrum of human knowledge, language, and culture (Gladkoff et al., 2023). Despite their capacity for generating new content, their output Is essentially a remix of existing human ideas and expressions. Thus, AI can be viewed as a digital manifestation of""Absent Presence"" where the human element, though not physically present, significantly influences the A''s output.

Recent research, including a study by Alemohammad et al. (2023), highlights the phenomenon of Model Autophagy Disorder (MAD), which underscores the dependence of AI on quality human input. MAD occurs when AI systems are trained repetitively on AI-generated content, leading to a decline in the quality and diversity of their output. This cycle risks creating a feedback loop that lacks the richness and variety inherent in original human contributions. The issue emphasizes the crucial role of human creativity in sustaining the effectiveness and relevance of AI models.

In the context of the development of emergent technologies and their integration into various aspects of society, it is essential to maintain a balance between leveraging their capabilities and acknowledging the indispensable role of human input. The concept of "Absent Presence" in AI invites a deeper exploration of the symbiotic relationship between human creativity and machine intelligence (Lucia, 2023). It serves as a reminder that while AI can significantly enhance our abilities to process and generate content, the essence of creativity and meaning-making remains a distinctly human endeavor (Gulya, 2023). Therefore, as AI continues to evolve and impact various facets of our lives, it is imperative to ensure that the human element remains central to its development and application. This approach will not only preserve the integrity and diversity of AI-generated content but also ensure that AI remains a tool that complements and extends human capabilities, rather than diminishing or replacing them.

Final considerations before moving forward include the profound impact of consciousness and cognition on language preservation and cultural identity. The activities engaged in, and the paradigms pondered, play a significant role in shaping the development of the brain itself and reinforcing neural pathways (Dang et al., 2023). This aspect becomes particularly critical in the context of educating children in environments dominated by a lingua franca, such as English. The pervasive use of a bridging language, especially in educational and technological contexts, can lead to a kind of cognitive and cultural intrusion. When children are educated primarily in a language other than their mother tongue, they may find themselves increasingly detached from traditional sources of meaning and cultural identity (Gu et al., 2023). This detachment is exacerbated by the omnipresence of digital devices and platforms, which often operate within a monolingual, predominantly English, framework. Such environments can act like a "walled garden" limiting exposure to diverse linguistic and cultural experiences (Kulko et al., 2023). Furthermore, relying predominantly on a bridging language like English can inadvertently lead to a divorce from other traditional paradigms of meaning-making, which are deeply embedded in different languages and cultures. Each language carries its unique semiotics – the study of signs and symbols and their use or interpretation – which is integral to the cultural and cognitive framework of its speakers (Madigan, 2023).

To counteract this, it is vital to intentionally create spaces and opportunities for engaging with and reinforcing native languages and cultural practices (Muhalim, 2023). This approach involves setting aside dedicated time and resources to foster and maintain linguistic and cultural connections. Such efforts are not only beneficial for current generations but are also a way of preserving these rich linguistic and cultural heritage for future generations. In essence, it is about striking a balance between embracing the utility of a lingua franca and maintaining the rich tapestry of diverse languages and cultures. This balance is

crucial for ensuring that while we leverage the benefits of a common language for global communication and technology, we do not lose the invaluable cultural and cognitive diversity that native languages embody. The goal is to ensure that future generations inherit a world rich in linguistic and cultural diversity, where they can act as avatars of their cultural heritage.

5. Recommendations for Future Studies

Building upon the proposed HAPPM in the context of Bujumbura, Burundi, the pathway forward for future studies in the preservation of cultural heritage is multifaceted and rich with potential. An essential step involves broadening the model's application across diverse cultural and geographical landscapes. By venturing into various regions, researchers can uncover the unique challenges and possibilities inherent in each, thereby refining HAPPM's adaptability and effectiveness. This expansion necessitates a comparative analysis of archival practices, offering a panoramic view of global preservation efforts and fostering a deeper understanding of cultural nuances.

In tandem with geographical diversification, the integration of emerging technologies represents a pivotal advancement for cultural preservation methodologies. The incorporation of artificial intelligence, blockchain, and virtual reality can revolutionize how we interact with and perceive cultural heritage, making it more accessible and engaging for a global audience. These technologies promise not only to enhance the immersive experience of cultural artifacts but also to ensure the integrity and longevity of digital archives. Central to the success of HAPPM and similar endeavors is the strength of interdisciplinary collaboration. The synergy between historians, data scientists, linguists, and technologists, enriched by insights from local communities and cultural experts, is crucial for navigating the complexities of digital humanities. Such collaborations ensure that preservation efforts are not only technologically sound but also culturally respectful and inclusive.

Addressing ethical considerations and data privacy emerges as a non-negotiable aspect of future research. As we delve deeper into the digital preservation of cultural artifacts, establishing robust frameworks for ethical research practices becomes paramount. These frameworks must navigate the delicate balance between accessibility and respect for the cultural significance and sensitivity of the data, ensuring that all preservation efforts are conducted with the utmost integrity. Moreover, enhancing public engagement and accessibility to preserved cultural heritage is essential for bridging the gap between archival repositories and the broader community. By creating interactive platforms and educational programs, researchers can ignite public interest and appreciation for cultural diversity, ensuring that the knowledge and beauty of our global heritage are shared widely.

As such, sustainability of digital preservation projects poses a significant challenge that future studies must address. Investigating innovative funding models and partnerships can secure the resources necessary for these projects' long-term success. Alongside financial sustainability, methodological evaluations and continuous improvement are vital, ensuring that our approaches to cultural preservation remain relevant, effective, and adaptable to the ever-evolving landscape of digital technology and cultural understanding. Through these concerted efforts, the field of cultural preservation stands on the cusp of a new era, poised to safeguard our global heritage for generations to come.

6. Conclusion

The preservation of ephemeral aspects of culture is now a critical imperative, whether these be grand cultural narratives or the seemingly mundane facets of everyday life. This preservation is increasingly vital in an era marked by the rapid evolution of global dynamics, where the unique richness of human experience, as expressed through diverse languages and cultural practices, faces the threat of being lost in the homogenizing tides of globalization and technological advancement. The loss of these linguistic and cultural nuances would represent an irreplaceable diminution of the human tapestry, as each language and cultural element distinctively contributes to the collective human knowledge and identity. The nuanced meanings and cultural contexts embedded in each language enrich our understanding of the world and foster deeper connections among people. The consideration of psychological implications, such as the alienation from on''s mother tongue and the cognitive effects of language loss, further emphasizes the need for sensitive and inclusive approaches to language preservation.

The methods and projects proposed offer a promising path forward with a comprehensive and scalable framework aimed at addressing the critical gaps in the existing paradigms of cultural preservation. By harnessing the capabilities of AI, big data, and machine learning, these approaches are designed to create a more nuanced and inclusive representation of cultural identities and individual stories. This marks a significant shift from traditional, often fragmented methods of preservation, towards a more integrated approach that captures the full spectrum of human experience. Thus, the urgency of language and cultural preservation stands out as a central theme. The rapid disappearance of languages across the globe highlights the immediate need for effective actions to document, revitalize, and sustain these vital components of cultural heritage. The innovative use of technology, particularly AI and computational linguistics, opens up new avenues for preserving languages and cultural nuances. These tools offer unprecedented capabilities for archiving and interpreting the vast array of human expressions, providing insights into both widely spoken and endangered languages.

In essence, the proposed projects and methodologies outlined in this article represent a holistic approach to cultural preservation, offering a comprehensive framework

to tackle the challenges of preserving not just the languages but also the everyday cultural practices that define human societies. By aggregating diverse archival data, these methods allow for a more complete and nuanced representation of cultural identities and individual personalities, moving beyond the limitations of traditional approaches. The next steps for research and implementation in this field are clear. There is a need for continued exploration and development of innovative methods to preserve linguistic and cultural diversity. This includes refining AI and computational linguistics tools, expanding archival projects, and fostering global collaborations. The goal is to ensure that the ephemeral, in all its forms, is not only remembered but also understood and appreciated by future generations. This endeavor is not just an academic or technological challenge but a moral imperative, crucial for maintaining the richness and diversity of our shared human heritage.

Declarations

Competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding

The current study was not financially supported by any organization/institutions.

Authors' contribution

The project's conceptualization was led by M. Ellsworth, with methodology development spearheaded by P. Ellsworth. Validation procedures were overseen by J. Hutson, who also conducted the investigation. The original draft was prepared by J. Hutson, who also handled the writing, review, and editing processes. Additionally, J. Hutson took charge of visualization tasks throughout the project.

Availability of data and materials

The presented articles can be found at the provided links.

Ethical considerations

The authors of this paper declare that all ethical issues have been meticulously considered and addressed prior to the submission of this manuscript for publication. This includes but is not limited to the avoidance of data fabrication and falsification, ensuring the originality of the work to prevent instances of double publication and submission, and eliminating redundancy. The integrity of the research has been upheld through the careful avoidance of plagiarism, with all sources appropriately cited and permissions obtained where necessary. Consent for publication has been secured, and any potential misconduct

has been thoroughly investigated and resolved.

References

Aalberse, S., Backus, A., & Muysken, P. (2019). *Heritage languages: A language contact approach*. John Benjamins Publishing Company. https://doi.org/10.1075/sibil.58

Ahlqvist, L., & Potter, B. B. (2023). Just a souvenir? Entangled identities within an early 2[0]th century American Indian basket collection. *Journal of Material Culture, 28*(3), 479-498. https://doi.org/10.1177/13591835221149685

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., & Baraniuk, R. G. (2023). Self-consuming generative models go mad. *37th Conference on Neural Information Processing Systems (NeurIPS 2023)* https://doi.org/10.52591/lxai202312101

Alexandrova, N. S. (2023). The Disappearance of Languages and Natural Bilingualism. *Polylinguality and Transcultural Practices, 20*(3), 436-455.

Anderson, R. (1999). Impact of first language loss on grammar in a bilingual child. *Communication Disorders Quarterly, 21*(1), 4-16. https://doi.org/10.1177/152574019902100102

Andleeb, N., Asgher, M., & Zimi, R. (2023). Exploration of speaking multiple language'' effect on cognitive flexibility and problem-solving skills. *Pakistan Journal of Humanities and Social Sciences, 11*(2), 837-845. https://doi.org/10.52131/pjhss.2023.1102.0394

Basu, A., Paul, S., Ghosh, S., Das, S., Chanda, B., Bhagvati, C., & Snasel, V. (2023). Digital Restoration of Cultural Heritage with Data-driven Computing: A Survey. *IEEE Access.*

Baugh, A., & Cable, T. (1993). *A history of the English language*. Routledge, London. https://doi.org/10.4324/9780203994634

Benson, C. (2010). How multilingual African contexts are pushing educational research and practice in new directions. *Language and Education, 24*(4), 323-336. https://doi.org/10.1080/09500781003678704

Bers, M. U. (2022). *Beyond coding: How children learn human values through programming.* MIT Press.

Blackledge, A., Creese, A., Arnaut, K., Blommaert, J., Rampton, B., & Spotti, M. (2018). Language and superdiversity. *The Routledge handbook of language and superdiversity*. Routledge, London. https://doi.org/10.4324/9781315696010-22

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution, 6*(2), 163-173. https://doi.org/10.1038/s41559-021-01604-y

Centeno, J. G. (2009, August). Issues and principles in service delivery to communicatively impaired minority bilingual adults in neurorehabilitation. *Seminars in Speech and Language, 30*(3): 139-152. https://doi.org/10.1055/s-0029-1225951

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology.*

Chen, G., Zhang, Y., Chen, N. S., & Fan, Z. (2023). Context-aware ubiquitous learning in science museum with ibeacon technology. In J. M., Spector, B. B., Lockee, & M. D. Childress, (Eds.), *Learning, design, and technology* (pp. 3369-3392). Springer, Cham. https://doi.org/10.1007/978-3-319-17461-7_5

Collette, V., & Kennedy, W. (2023). *A concise dictionary of Nakoda (Assiniboine).* University of Nebraska Press. https://doi.org/10.2307/j.ctv360ns3q

Cruz, M. (2024). Exploring the integration of artificial intelligence generative tools in teaching Hispano-American literature: A student-centric approach. In *INTED2024 Proceedings* (pp. 5717-5727). IATED.

Crystal, D. (2003). *English as a global language*. Cambridge university press.

Dang, Q., Ma, F., Yuan, Q., Fu, Y., Chen, K., Zhang, Z., Lu, C., & Guo, T. (2023). Processing negative emotion in two languages of bilinguals: Accommodation and assimilation of the neural pathways based on a meta-analysis. *Cerebral Cortex, 33*(13), 8352–8367. https://doi.org/10.1093/cercor/bhad121

Diamond, J. (2010). The benefits of multilingualism. *Science, 330*(6002), 332-333. https://doi.org/10.1126/science.1195067

Dildora, U., & Tukhtakhodjayeva, Z. T. (2024). The socio-cultural characteristics of formulaic language and Uzbeck languages. *Journal of*

new century innovations, 48(2), 21-23.

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). Making connections between theory and practice. Motivation and foreign language learning: From theory to practice *Language Learning & Language Teaching 40.* (pp. 177-183). https://doi.org/10.1075/lllt.40.09doi

El-Gabalawi, F., & Khouzam, H. (2009). The effect of mental illness on language regression to the mother tongue in bilingual teenagers. *The Journal of Neuropsychiatry and Clinical Neurosciences, 21*(1), 88-91. https://doi.org/10.1176/jnp.2009.21.1.88

Gallo, F., & Abutalebi, J. (2023). The unique role of bilingualism among cognitive reserve-enhancing factors. *Bilingualism: Language and Cognition,* 1-8. https://doi.org/10.1017/S1366728923000743

Gannie, R. R. (2023). Destruction and resiliency: Decolonizing settler knowledge in native 137ersonaln literature through the peoplehood matrix [Doctoral dissertation, University of Denver, United States].

Gladkoff, S., Erofeev, G., Han, L., & Nenadic, G. (2023). Predicting Perfect Quality Segments in MT Output with Fine-Tuned OpenAI LLM: Is it possible to capture editing distance patterns from historical data?. arXiv preprint arXiv:2308.00158. https://arxiv.org/abs/2308.00158

Govia, L. (2020). Coproduction, ethics and artificial intelligence: A perspective from cultural anthropology. *Journal of Digital Social Research, 2*(3), 42-64. https://doi.org/10.33621/jdsr.v2i3.53

Graddol, D. (1997). *The future of English?: A guide to forecasting the popularity of the English language in the 21st century*. British Council.

Grohmann, K. K., Kambanaros, M., Leivada, E., & Rowe, C. (2016). A developmental approach to diglossia: bilectalism on a gradient scale of linguality. Poznan Studies in Contemporary Linguistics, 52(4), 629-662. https://doi.org/10.1515/psicl-2016-0025

Gu, M. M., Chiu, M. M., & Li, J. Z. (2023). Adult ethnic minoritie'' mainstream language proficiency: cultural knowledge, cultural identification, and language use attitudes. Language, Culture and Curriculum, 1-21. https://doi.org/10.1080/07908318.2023.2210287

Gulya, J. (2023). Rebranding Originality for the Age of AI. *International Journal of Emerging and Disruptive Innovation in Education: VISIONARIUM, 1*(1), 5.

Hale, S. A. (2015, April). Cross-language Wikipedia editing of 137 ersonaa, 137 erson. In Proceedings of the 3³rd Annual ACM Conference on Human Factors in Computing Systems (pp. 183-192). https://doi.org/10.1145/2702123.2702346

Hinkel, E. (2001). Why English passive is difficult to teach (and learn). *New perspectives on grammar teaching in second language classrooms* (pp. 245-270). Routledge. https://doi.org/10.4324/9781410605030-20

Hollebeke, I., Dekeyser, G. N., Caira, T., Agirdag, O., & Struys, E. (2023). Cherishing the heritage language: Predictors of parental heritage language maintenance efforts. *International Journal of Bilingualism, 27*(6), 925-941.

Hutson, J., Ellsworth, P., Ellsworth, M., & Ntungirimana, J. B. (2023). Digitizing the Cultural Capital: Harnessing Digital Humanities for Heritage Preservation in Bujumbura, Burundi. *Advances in Social Sciences and Management, 1*(11).

Jerome, C., Perry, E. J., & Ting, S. H. (2022). Malaysian pop music in the minority language: A tool for forging unity and a common identity?. *International Journal of Social Science Research, 4*(1), 401-413. https://myjms.mohe.gov.my/index.php/ijssr/article/view/17731

Kandler, A., & Unger, R. (2023). Modeling language shift. In A. Bunde, J. Caro, C. Chmelik, J. Kärger, & G. Vogl (Eds.), *Diffusive spreading in nature, technology and society* (pp. 365-387). Springer. Cham. https://doi.org/10.1007/978-3-031-05946-9_18

Kazanas, S. A., & Altarriba, J. (2012). Multilingualism and memory. *The Encyclopedia of Applied Linguistics* Wiley (pp. 1-7). https://doi.org/10.1002/9781405198431.wbeal1448

Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2023). Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with137ersonalizedd feedback. *arXiv preprint arXiv:2303.05453.*

Koch, M. J., Greve, W., & Kersten, K. (2023). Extending the cognitive advantage hypothesis: a conceptual replication study of the relationship between multilingualism and flexible goal adjustment. *Journal of Multilingual and Multicultural Development*, 1-15. https://doi.org/10.1080/01434632.2023.2189262

Konieczny, P., & Lewoniewski, W. (2023). Measuring Americanization: A global quantitative study of interest in American topics on Wikipedia. arXiv preprint arXiv:2307.14401. https://arxiv.org/abs/2307.14401

Kroch, A., Taylor, A., & Ringe, D. (2000). The Middle English verb-second

constraint. Textual Parameters in Older Languages, 353-392. https://doi.org/10.1075/cilt.195.17kro

Kulko, R. D., Pletl, A., Mempel, H., Wahl, F., & Elser, B. (2023). OpenVNT: An open platform for VIS-NIR technology. *Sensors, 23*(6), 3151. https://doi.org/10.3390/s23063151

Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. S. (2019). What influences language impairment in bilingual aphasia? A meta-analytic review. *Frontiers in Psychology, 10,* 445. https://doi.org/10.3389/fpsyg.2019.00445

Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). Chat GPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. arXiv preprint arXiv:2304.05613. https://doi.org/10.18653/v1/2023.findings-emnlp.878

Lanza, E., & Svendsen, B. A. (2007). Tell me who your friends are and I might be able to tell you what language (s) you speak: Social network analysis, multilingualism, and identity. *International Journal of Bilingualism, 11*(3), 275-300. https://doi.org/10.1177/13670069070110030201

Lucia, A. (2023). Flooding the Web: Absence-presence and the media strategies of Nithyananda's digital empire. *Gurus and media: Sound, image, machine, text and the digital,* 271-298. https://library.oapen.org/bitstream/handle/20.500.12657/76564/1/9781800085541.pdf#page=290

MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal, 54*(1), 167-201. https://doi.org/10.3102/0002831216683935

Madigan, T. (2023). Cultural syntax and the rules of meaning-making: A new paradigm for the interpretation of culture. In A. Cossu & J. Fontdevila (Eds.), *Interpretive sociology and the semiotic imagination* (pp. 118-139). Bristol University Press. https://doi.org/10.51952/9781529211764.ch005

Makhmudov, K. (2023). Bridging cultures through English language education: a comprehensive model for intercultural communication competence development. *ISJ Theoretical & Applied Science, 3*(119), 204-208. https://doi.org/10.15863/TAS.2023.03.119.27

Martin, C., Woods, B., & Williams, S. (2019). Language and culture in the caregiving of people with dementia in care homes-what are the implications for well-being? A scoping review with a Welsh perspective. *Journal of cross-cultural gerontology, 34*, 67-114. https://doi.org/10.1007/s10823-018-9361-9

Morgenstern, A., & Boutet, D. (2024). The 0rchestration of bodies and artifacts in French family dinners. In *Diachronic Perspectives on Embodiment and Technology* (pp. 111-130). Springer, Cham.

Muhalim, M. (2023). Negotiating religious discourses in English language teaching: Reorienting and reframing dominant English ideologies. *Changing English, 30*(3), 209-222. https://doi.org/10.1080/1358684X.2023.2217424

Myers-Scotton, C. (2003). Code-switching: Evidence of both flexibility and rigidity in language. In J. M. Dewaele, A. Housen, &W. Li (Eds.), *Multilingual Matters* (pp. 189-203). Multilingual Matters, Bristol. https://doi.org/10.21832/9781853596315-014

Ng, B. C., & Wigglesworth, G. (2007). *Bilingualism: An advanced resource book*. Taylor & Francis.

Obler, L. K., Park, Y., Gitterman, M. R., & Goral, M. (2012). The study of bilingual aphasia: The questions addressed. In M. R. Gitterman, M. Goral, & L. K. Obler, *Aspects of Multilingual Aphasia* (pp. 3-15). Multilingual Matters, Bristol. https://doi.org/10.21832/9781847697554-003

Ochieng, V. O., & Waithanji Ngware, M. (2023). Adoption of education technologies for learning during COVID-19 pandemic: The experiences of marginalized and vulnerable learner populations in Kenya. *International Journal of Educational Reform, 32*(4), 464-487. https://doi.org/10.1177/10567879221076081

Park, S. (2023). Multilingualism, social inequality, and the need for a universal language. *Journal of Universal Language, 24*(1), 77-93. https://doi.org/10.22425/jul.2023.24.1.77

Parkinson, B. (2023). Cross-Cultural Calibration of Words and Emotions: Referential, Constructionist, and Pragmatic Perspectives. *Emotion Review, 15*(4), 348-362. https://doi.org/10.1177/17540739231182680

Phillipson, R. (1992). *Linguistic imperialism*. Oxford University Press.

Ramesh, K., Sitaram, S., & Choudhury, M. (2023). Fairness in language models beyond English: Gaps and challenges. arXiv preprint arXiv:2302.12578. https://doi.org/10.18653/v1/2023.findings-eacl.157

Rataj, K., Kakuba, P., Mandera, P., & van Heuven, W. J. (2023). Establishing

semantic relatedness through ratings, reaction times, and semantic vectors: A database in Polish. *Plos One, 18*(4), e0284801. https://doi.org/10.1371/journal.pone.0284801

Salomone, R., & Salomone, R. C. (2022*). The rise of English: Global politics and the power of language.* Oxford University Press.

Savaya, R., Berger, R., Ronen, T., & Roziner, I. (2023). Social alienation and psychological distress in 138sraeli adolescents. *Child and Adolescent Social Work Journal, 40*(3), 395-408. https://doi.org/10.1007/s10560-021-00786-5

Sefotho, M., Charamba, E., & Quintero, G. (2023). Translingualism across languages: A textual analysis of languages interaction. *Education Innovation Diversity, 1*(6), 16-30. https://doi.org/10.17770/eid2023.1.6968

Sevinç, Y. (2022). Mindsets and family language pressure: language or anxiety transmission across generations?. *Journal of Multilingual and Multicultural Development, 43*(9), 874-890. https://doi.org/10.1080/01434632.2022.2038614

Sharma, R., Morwal, S., & Agarwal, B. (2021). Entity-extraction using hybrid deep-learning approach for Hindi text. *International Journal of Cognitive Informatics and Natural Intelligence, 15*(3), 1-11. https://doi.org/10.4018/IJCINI.20210701.oa1

Shcherbakova, O., Michaelis, S. M., Haynie, H. J., Passmore, S., Gast, V., Gray, R. D., Greenhill S. J., Blasi, D., & Skirgård, H. (2023). Societies of strangers do not speak grammatically simpler languages. *Science Advances*, 1-37. https://doi.org/10.31235/osf.io/svfdx

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., ... & Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. Science Advances, 9(16), 6175. https://doi.org/10.1126/sciadv.adg6175

Skutnabb-Kangas, T. (2013). *Linguistic genocide in education--or worldwide diversity and human rights?.* Routledge.

Szubko-Sitarek, W. (2015). *Multilingual lexical recognition in the mental lexicon of third language users.* Heidelberg: Springer.

Tankosić, A., & Dovchin, S. (2023). The impact of social media in the sociolinguistic practices of the peripheral post-socialist contexts. *International Journal of Multilingualism, 20*(3), 869-890. https://doi.org/10.1080/14790718.2021.1917582

Torres-Zúñiga, L., & Söğüt, S. (2024). Using Intercultural Virtual Exchange to Promote Critical Pedagogy Practices of English Language Teachers. *International Perspectives on Critical English Language Teacher Education: Theory and Practice*, 217.

Van der Worp, K., Cenoz, J., & Gorter, D. (2017). From bilingualism to multilingualism in the workplace: The case of the Basque autonomous community. *Language Policy, 16*, 407-432. https://doi.org/10.1007/s10993-016-9412-4

Wang, P. J. (2024). Exploring AI's role in bilingual Bible education: Overcoming language barriers- A case study of sixth grade bible story reading at Sankuang Elementary School in New Taipei City. *In INTED2024 Proceedings* (pp. 1671-1680). IATED.

Wei, L. (2014). Translanguaging knowledge and identity in complementary classrooms for multilingual minority ethnic children. *Classroom Discourse, 5*(2), 158-175. https://doi.org/10.1080/19463014.2014.893896

Zhang, M., Jiang, G., Liu, S., Chen, J., & Zhang, M. (2024). LLM–Assisted Data Augmentation for Chinese Dialogue–Level Dependency Parsing. *Computational Linguistics,* 1-24.

Zhang, J., Ji, X., Zhao, Z., Hei, X., & Choo, K. K. R. (2023). Ethical considerations and policy implications for large language models: guiding responsible development and deployment. *arXiv preprint arXiv:2308.02678.*

Zhang, A., & Zhu, X. (2023). Analysis of English translation of corpus based on blockchain. *International Journal of Web-Based Learning and Teaching Technologies, 18*(2), 1-14. https://doi.org/10.4018/IJWLTT.332767