Language revitalization is an urgent mission in the field of linguistics, and there are ways in which language technology can help. According to the United Nations, “one indigenous language dies every two weeks” (The United Nations Permanent Forum on Indigenous Issues [UNPFII], 2018). This is largely due to colonization, which has significantly harmed and suppressed indigenous communities and their languages. There is a great lack of resources about these languages, and when resources do exist, they often do not meet the needs of the community (Meighan, 2021). Colonization has also caused immense decreases in indigenous populations, leading to over half of the world’s languages being spoken by a very small percentage of the global population (UNPFII, 2018). This disparity highlights the urgency of language revitalization, and why indigenous communities and linguists may turn to language technology for help. The following pie chart shows the United Nations’ estimates of these numbers.

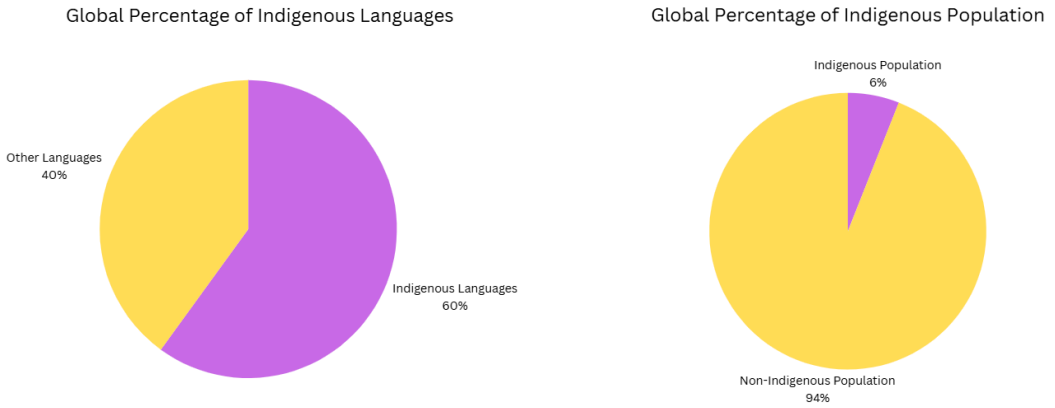

Figure 1. Highlights the drastic difference between the proportion of languages that are indigenous and the proportion of the population that speak those languages. Data collected from The United Nations Permanent Forum on Indigenous Issues, 2018. Graphic created using Canva.

Linguists and researchers have made efforts to incorporate language technology into the process of language revitalization. Two main forms of language technology used for this work are Automatic Speech Recognition (ASR), also called Speech to Text, and Machine Translation (MT) (Viannis, 2024).

The first step in the process of language revitalization is data gathering and documentation. Many endangered languages don’t have a writing system, which makes ASR particularly useful. It can be used to convert speech into written text, creating data for linguists to analyze. Another benefit of using ASR is that it requires the direct involvement of native speakers, which improves the accuracy of the data and gives the speakers leadership in the process of preserving their language (Kavitha et al., 2023).

After data collection, linguistic experts are needed to create accurate and accessible learning materials, which is where MT comes in. MT allows the transfer of knowledge that only native speakers

have into a language that is more widely accessible. This allows field linguists to discover patterns and grammatical structures in the language, as well as variations in dialects (Kavitha et al., 2023). With this information, linguists are able to create language learning resources tailored to specific communities. Combining this linguistics expertise with language technology can also lead to resources such as language learning applications and chatbots (Kavitha et al., 2023).

Even with these incredible technologies, there are still various challenges in language revitalization. For example, "resource limitations, dialectal variations, [and] ethical considerations" are some of the most prominent hurdles in using language technology for revitalization (Kavitha et al., 2023). Many endangered languages lack the substantial corpora needed to train language models effectively (Kavitha et al., 2023; Viannis, 2024). ASR can provide more data, but that is typically not a simple task. Many communities don't have internet access or other infrastructure needed to use these technologies, so the challenge of sparse data remains (Viannis, 2024). Another important drawback of using language technology for language revitalization is that language models prefer standardized linguistic patterns, which can lead to the loss of dialects within an endangered language, especially if data is scarce (Viannis, 2024).

Aside from the challenges related to the technology itself, there are also ethical concerns that must be considered in the process of language revitalization. The involvement of various parties (technology experts and companies, linguists, native speakers, etc.) raises the question of ownership of the data produced and collected (Kavitha et al., 2023). An example of the issues that can arise from this is when the Standing Rock Sioux tribe worked with the Lakota Language Consortium to preserve their native language. When a member of the tribe, Ray Taken Alive, "asked for copies [of the materials], he was shocked to learn that the consortium, run by a white man, had copyrighted the language materials, which were based on generations of Lakota tradition. The traditional knowledge gathered from the tribe was now being sold back to it in the form of textbooks" (Brewer, 2022). In addition to avoiding situations like this, it is vital to maintain faithful cultural representations of the speakers and ensure that language models "do not inadvertently exploit or misrepresent endangered language communities and their heritage" (Kavitha et al., 2023).

The involvement of language technology can be harmful to indigenous communities, so communication is crucial in order to determine if involving technology is the right choice (Arppe et al., 2023). Micheal Running Wolf, an indigenous software engineer, works with indigenous communities to create language technologies for them. For one community, he explains that they generated "'500 phrases in Makah and Kwak'wala, defined by the community and also the rules of the language, obviously, and [they] trained the AI to recognize those 500 phrases and those 500 phrases are used in curricula'" (Maracle, 2024). Running Wolf also highlights the importance of the indigenous community being in

charge of this process, stating that "'we need to have sovereignty over our own data, set the terms and that's the only way to build this AI.'" In some cases, taking legal precautions is the best approach. (Maracle, 2024). Collaboration like this can lead to tailored solutions that meet the unique needs of each language community, and involving native speakers in the process of designing uses for language technologies ensures "long-term sustainability of language preservation efforts" (Viannis, 2024).

Although it has drawbacks, language technology can aid in the process of language revitalization. ASR and MT allow for indigenous languages to become more accessible to field linguists, leading to resources that can better help communities revitalize their language. The decision of whether or not to use these technologies is ultimately up to the indigenous community, and with collaboration and communication from all parties, they can ensure that they receive help and resources that meet the community's needs.

References:

Arppe, A., Hermes, M., Junker, M. O., Livesay, N., & Running Wolf, M. (2023). Algonquian Conference. *Computational Linguistics, Language Technologies and the International Decade of Indigenous Languages: Academic and Community Interactions*. Boulder. Retrieved from https://celj.cu.law/?p=935.

Brewer, G. L. (2022, June 3). *Lakota elders helped a white man preserve their language. then he tried to sell it back to them.* NBCNews.com. https://www.nbcnews.com/news/us-news/native-american-language-preservation-rcna31396

Kavitha, R., Sherin Fathma, S., & Vinitha, V. (2023). AI: A Catalyst for Language Preservation in the Digital Era. *Shanlax International Journal of English*, *12*(1), 102–106. https://doi.org/10.34293/rtdh.v12is1-dec.39

Maracle, C. (2024, August 12). *How AI can help indigenous language revitalization, and why data sovereignty is important*. CBCnews. https://www.cbc.ca/news/indigenous/ai-indigenous-languages-1.7290740

Meighan, P. J. (2021). Decolonizing the digital landscape: The role of technology in indigenous language revitalization. *AlterNative: An International Journal of Indigenous Peoples*, *17*(3), 397–405. https://doi.org/10.1177/11771801211037672

United Nations. (2018). *Indigenous Languages Language Rights of Indigenous Peoples*. UN Department of Public Information. https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/04/Indigenous-Languages.pdf

Viannis, O. (2024, December 20). *AI-Powered Preservation of Endangered Languages*. Historica.org; Historica. https://www.historica.org/blog/ai-powered-preservation-of-endangered-languages