Technology's Dual Role in Language Marginalization and Revitalization

Thomas Yim

Stanford University

Introduction

When Oxlaj Kumez goes onto the internet or turns on the television, his language is nowhere to be found. Sending texts, writing emails, reading news articles, and scrolling through social media have become ubiquitous activities in the modern digital age. But these services and tools are simply unavailable to Kumez in Kaqchikel, an indigenous Mayan Language. When asked in an interview for a BBC article about his experience using digital technology, he commented: "Why am I going to teach it to my children if, when I turn on the internet or television, I cannot find it there?" This is a shared experience among many speakers whose minority languages have not been introduced to the digital world. Among the thousands of existing languages, only a few hundred have an internet presence and the most common ten languages account for over three-fourths of the language content online (Statista, 2020).

Many scholars have responded to experiences like Kumez's and the general inaccessibility of most languages online, criticizing the monumental language shift towards globally dominant languages such as English (Bender, 2011; Cunliffe & Herring, 2005; Kornai, 2013). But others highlight how digitalization allows some communities to stay connected over long distances or store language documentation in an accessible form online (Meighan, 2021; Leonard, 2023). According to an estimate by the United Nations Educational, Scientific and Cultural Organization, 43% of languages used in the world are endangered (UNESCO, 2010, p.4). Largely the result of language persecution through massacres, such as "La Matanza," and educational colonialism in government boarding schools, many indigenous languages have few speakers left. And more recently, technology has been responsible for a different form of language marginalization. The transition to digital communication has heavily incentivized languages like English, Chinese, and Spanish, which already dominate online interaction.

When discussing "language marginalization," I refer to the social pressure as Lippi-Green describes as "any speaker of a stigmatized vernacular is promised large returns if he or she will adopt [Standard American English]." (Lippi-Green, 1990, p. 71). Whereas her argument references how certain dialects are treated as the standard and superior form, it also applies to languages as a whole and how society incentivizes language switching, unless a speaker wishes to be "cut off from everyday privileges and rights," including technology (Lippi-Green, 1990, p.

71). Furthermore, I use “minority languages” and “indigenous languages” interchangeably as it is often the case that indigenous languages are endangered and have a minority share of speakers. Communities that use these languages also call them “treasure” languages, to emphasize their continued existence and centrality to their lives.

So why has the internet had such an adverse effect on language diversity? And what can we do to change that? Initially, I envisioned a technological solution. Several years ago I wrote a Python program applying phonological rules of vowel stress to the dictionary of the Kashaya Language, the critically endangered language of the Kashia band of the Pomo people. Quickly identifying many entries in the dictionaries that deviated from the rules that were originally documented in the language, computational acceleration deeply impressed me at first. Since then, my perspective has shifted as I discovered the complexity of the relationship between language documentation, linguistic data sovereignty, and digital technology – especially given how minority languages have been marginalized by the English-dominant internet. In this paper, I examine how technology has marginalized indigenous languages and present a case for language diversity.

First, to explore how content is produced online and how it has shifted from when the internet was created, I discuss the importance of encodings in making a variety of scripts and digital characters available. Second, to understand how technology incentivizes majority language usage but also aids in revitalization programs, I review literature from socio and computational linguists such as Emily M. Bender, Daniel Cunliffe, Susan Herring, András Kornai, and Jenny Davis. Third, I explore the perspectives of the affected language communities to better understand whether these technological resources actually help with current revitalization efforts. To do so, I interview Wesley Leonard, a citizen of the Miami Tribe of Oklahoma and a leading scholar in the discussion of indigenous data sovereignty and language reclamation, about his views on digital language use. Indigenous data sovereignty is the right of indigenous people to have control over the storage and usage of data related to their identity, including language. For Leonard, there should be no controversy over a researcher’s rights to language data versus indigenous ownership – indigenous data sovereignty “just is,” (W. Leonard, personal communication, February 13, 2023). Finally, I analyze the effectiveness of the computer vision translation apps Google Woolaroo and Kupu through a lens of community-driven development. I also interview Chance Coughenour, one of the creators of Woolaroo, to study

how Google developed and built the app. Digital technology presents novel forms of representing and storing languages and can play a critical role in documentation and relearning programs.

So even though the digitalization of language and the rise in internet communication has largely been responsible for marginalizing minority languages, that same technology is equally important as a tool for documenting, reclaiming, and learning these endangered languages. However, for digital revitalization to be effective, heavy consideration must be given to indigenous data sovereignty and indigenous communities' incentives to have any hope of the project gaining traction or widespread use in the language.

This paper aims to add to the literature and discussion on online minority-majority language interaction that tends to ignore the community impact and perspective. I propose a framework to guide working with indigenous language data and then use it to analyze how computational technology can influence language reclamation efforts and the communities they belong to. Specifically, the researcher or developer must have explicit consent for the collection and usage of all language data, make the data and findings accessible and removable by the original owners of that data, and be in ongoing communication about the intended use case and goals of the project (see Figure 1).

Researcher positionality

How a researcher involves the community can be as important – if not more – as the content of the research itself. As an L1 speaker of English, I am not a member of the indigenous language communities highlighted in this paper. I do not aim to speak for them nor generalize their opinions but merely aim to combine the literature of past scholars and Dr. Leonard's perspective to analyze how technology affects reclamation efforts. As a former computational linguistics researcher, I used to perceive technology as only having a beneficial impact on language revitalization because of the automation, efficiency, and connectedness it offers, but I have since shifted my thinking to consider the dimensionality and lack of clear delineation between help and harm.

Digitalization of language

Since English has been the primary online language and programming languages use Latin scripts, the digital era has created an even more Anglo-centric world where people must assimilate to English or other majority languages to take advantage of all the efficiency it offers. But it doesn't have to be that way. Computers and internet protocols transmit and store data through binary, a series of ones and zeros, and it is only through its encoding system that computers convert those numbers to letters. These digital tools have just been created and better developed in English over time which has led to its dominant usage online.

How a computer converts its binary and displays digital characters depends on its encoding system. For over thirty years, ASCII was the only encoding for letters and was limited to just 128 characters. This was enough to represent any Latin script but was inaccessible to language communities using anything else. Other than images, audio recordings, or videos, there was no way to digitally represent a language whose characters were excluded by ASCII. Languages like English, Spanish, and French could mostly be produced online, but Arabic, Cherokee, and Korean could not. Speakers of these other languages were forced to "[develop] unofficial phonetic representations of their language using standard roman characters," and assimilate to the systems and standards at their disposal (Cunliffe & Herring, 2005, p.131). And as the usage of computers and the internet increased, it became a major inconvenience to not be able to use one's native script on the many apps and digital services that were developed. When people cannot find and use their language online, there is increased use of English both in consumption and production which contributes heavily to a language shift (Cunliffe & Herring, 2005). Since these tools would not accommodate their languages, the languages had to adapt.

However, there have been movements and projects to make technology more accessible for a variety of languages. The introduction of a new encoding system, Unicode, increased the number of displayable letters from 128 to over 1.1 million (Unicode Standard, n.d.). At the time this paper was written, computers were using Unicode version 15.0 which contains 149,186 characters – meaning that the scripts of many other languages could be incorporated into the standard. The Unicode Standard states that its goal was to "support the needs of all types of users,… using mainstream or minority scripts" (Unicode Standard, n.d).

Even though other languages are becoming more accessible and visible online, English or Romanized character representations of other languages continue to dominate, having accumulated enough momentum for having been the only option for so long. Nomin-Erdene Bayarsaikhan, a student at Stanford University, grew up in Mongolia but often uses English when texting with other Mongolian speakers. She attributes this to a variety of factors including familiarity with the English keyboard, continuing to use English by habit, and the intimacy and informality they associate with English characters (N. Bayarsaikhan, personal communication, March 12, 2023). Regardless of how it has persisted, English still has a significant influence over this form of communication. Just as many languages borrow words from each other, there has been an adoption of language that has become the norm in digital usage.

Just because more diverse language content exists and has become accessible online does not mean that its users will read or interact with it. Computational linguist, Andras Kornai, in his paper "Digital Language Death," has a more pessimistic view of the effectiveness of digital resources in promoting minority language usage. He argues that "online audio files of an elder tribesman reciting folk poetry will not facilitate digital ascent" (Kornai, 2013, p. 2.). Digital ascent is the ability of a language to achieve regular usage and interaction online. Using statistical regression models considering features such as the number of speakers or the age of the speakers, he concludes that "the vast majority (over 95%) of languages have already lost the capacity to ascend digitally" (Kornai, 2013, p. 2). This loss suggests that even if we produce documentation, and minority languages become more accessible online, there is no guarantee that the language communities are going to use it or find it helpful. To facilitate strong online interaction, there must be two-way channels of communication, such as email and social media.

Thus, simply having dictionaries or translated texts in the native language may have little impact on its online usage. In Dr. Leonard's language community, people have tried to build apps or digital games to facilitate engagement with the language, but he argues that they are rarely as effective as just being in a situation where people are using the language (W. Leonard, personal communication, February 13, 2023). So, regardless of the intent or how technologically advanced a resource is, I argue that a lack of community-centered design and ideation often renders these projects useless.

So, the question shifts toward how we can change technology to promote minority language usage. Jenny Davis, a citizen of the Chickasaw Nation and professor of sociolinguistics, has a more optimistic view that languages, as exemplified by Wendat, Miami, and Wampanoag can re-awaken even if there are no speakers left (Davis, 2017). She draws a distinction between dormancy and extinction where the finality of language endangerment is removed. And it is through proper documentation and preservation of these language materials that a language could be relearned.

Now that the various technological barriers have been removed, what is the next step in solving this issue and promoting minority language use? What resources should developers build that will actually have the greatest impact and usage? The answer can be found through constant communication with indigenous communities about what projects would be most beneficial.

Misaligned incentives for outsiders' support in language reclamation efforts

There is often a lack of trust from indigenous language communities in linguistic researchers and their incentives for collaboration. They worry about public access to language data and how non-members of the community often use research or the data to make false claims to Indigeneity for personal gain (Shulist & Pedri-Spade, 2022). For example, in the Anishinaabe community, a settler was hired to aid in revitalization efforts as a teacher. But when his teaching methods were questioned, he claimed to know the language and culture better than many others in the community. He then attempted to use this rhetoric to justify his teaching methods. This was seen as a form of cultural imperialism and a false claim to their indigenous identity. Linguists conducting documentation of languages often use these languages to advance their academic careers without properly crediting the people it is sourced from. So, while language revitalization research is often beneficial, it can be a dangerous situation where the scholar not only adopts the language itself, but also "cultural meanings and practices, and possibly over Indigenous identity itself" (Shulist & Pedri-Spade, 2022, p. 273).

Furthermore, there are issues that arise from misaligned incentives and goals. Often, these apps are built without considering how the language is really used or how the community wants the language to be used. Researchers have a "tendency… to equate language with data, and by extension to view language as a mere data point… which serves to dehumanize and

decontextualize [it]" (Holton et al., 2022, p. 50). There needs to be increased emphasis on ongoing community consent, involvement, and conversation throughout the entire development process. If a researcher is praised for helping these indigenous communities by building certain resources, but no one from the indigenous community needs it or wants to use it, they are taking advantage of the endangered status of the language and are partaking in a modern form of linguistic colonialism. Endangered language research has enabled academics, "to treat Indigenous languages as 'things'… and entail the potential for people to pick them up and re-use them for their own gain" (Shulist & Pedri-Spade, 2022, p. 281). These researchers are considered saviors of these languages and are praised for helping preserve them – regardless of if anyone uses those materials. In academics, it is often the case that "primary and structural data may be construed by some as research products and thus creations of the researcher rather than the language community" (Holton et al., 2022, p. 50). Thus, when building dictionaries, translation apps, language learning materials, and other digital content, it is paramount to understand its purpose and acknowledge who the data belongs to.

Cooperation and clarification of goals

Decisions about linguistic data usage and sovereignty must come from the community itself. There has been research and increasing discussion about indigenous peoples' rights to control the collection, storage, and usage of their cultural data (Noone, 2022). The Māori community has discussed how "by digitizing Māori data and information without permission or consultation, [researchers] have breached traditional Māori customary rights and beliefs" (Taiuru, 2018, p. 3). When that trust in the researcher and their project disappears, it is unlikely that anyone will use it. So, regardless of how well-built these digital language tools are, Kornai is correct in his judgment that there are no guarantees of their impact. I continue my exploration of responsible research practices in an interview with Dr. Wesley Leonard, during which I compiled three main principles of building resources that involve the communities themselves.

First, there must be explicit consent over all collection and usage of language data. Any time a researcher acquires indigenous language data, the person or community who is the source of that data must give full informed consent to all aspects of the project. Furthermore, all subsequent research that wishes to use that data must get consent for its specific application. The

longstanding negligence to obtain consent for English content is in machine learning and natural language processing research because the creators allegedly shared it online knowing it would be public is finally gaining attention (Frenkel &Thompson. 2023). Such unconsented collection in English should not be carried over to indigenous language resources. Similarly, legacy data should not have the assumption of open usage as its original consent may not apply to more modern advances in technology.

Second, data that has been collected must be accessible to anyone from that language community, and they have the right to remove any of it. For example, in Leonard's tribe, they agreed to make the dictionary public but removed all proper nouns out of privacy. Ray Taken Alive, a member of the Lakota people and teacher of the language, expressed that "no matter how it was collected, where it was collected, when it was collected, our language belongs to us" (NBC, 2022). Wilhelm Meya, who is not a member of the Lakota community, started the Lakota Language Consortium and documented the language (Brewer, 2022). But he copyrighted the data and asked the Lakota people to pay for the textbooks. Meya's actions exemplify the common misalignment of profits and goals from these projects, and these claims over data ownership eradicate trust in these research relationships.

Finally, in the development of these resources, the outside party and indigenous community must be in constant, ongoing communication about what they want from it. Rather than just have solutions be prescribed to minority language communities, the developer of the tool should clarify what features the community would find useful. The following section is an analysis of Google Woolaroo, an app that translates pictures into words in a variety of indigenous languages (see Figure 2), through the framework of these three principles and a lens of community-driven development.

A case study of Google Woolaroo

Woolaroo was created by Google Arts and Culture, which is a nonprofit organization. Any user can take a picture and get the word of the object translated into Yugambeh, Māori, Nawat, and many other languages. I conducted an interview with Chance Coughenour, a member of the team that built Woolaroo, during which he shared how Google approached the development process of the app.

Google made the code open source to demonstrate full transparency of their methods and only uses the data that the communities provide to them (C. Coughenour, personal communication, February 22, 2023). The practice of open source guarantees accountability, which should be standard practice in linguistic research. Google also pays for the app to be hosted online so it is freely accessible to anyone. But the language community could deploy their own version with that code if they wanted even more control. Furthermore, the providers of the language data retain full legal ownership and have an uncontested right to remove it at any time (C. Coughenour, personal communication, February 22, 2023). The degree of control given to these communities shows a dedication to usage consent and data accessibility and removability. Especially with how multinational corporations like Google often have a bad reputation with smaller communities, this openness is crucial to building trust with communities.

Furthermore, Rory O'Connor, CEO of the Yugambeh Museum, was a part of the creation of the app and had influence over its features. He wrote that the app "is open source and allows language communities like ours to preserve and expand their language word lists and add audio recordings to help with pronunciation" (O'Connor, 2021). Building the app with his input guaranteed usage because it is truly fitting some need. O'Connor further proposed that they add a feature where anyone can add a word and make the product even more open source. However, this open contribution policy could fall under the linguistic colonialism detailed by Shulist and Pedri-Spade. Being completely open and editable by anyone means that non-members can contribute which risks malpractice and artificial claims to ownership of that language data and indigeneity. Such a potential for misuse raises the need for methods of authentication and security for the data against bad actors – something Woolaroo has not implemented yet.

So far, while the lack of security presents a potential ownership issue, no one has ever added a word. There has not been widespread adoption of this tool and Chance even admits that despite Woolaroo's promising applications, it is not being used as frequently as he would have hoped. A similar visual translation app, called Kupu was developed to help teach Māori in the classroom (see figure 3). With 177,000 downloads, 2.7 million image translations, and 5 million word pronunciations played, it has become a popular way of learning and translating Māori (Keall, 2019). They recently released a desktop and tablet version after overwhelming requests from Māori teachers nationwide who wanted the app to be accessible on various devices for their students (Keall, 2019). This dedication to adapting the project to the needs of its actual users

explains its success and has compounded its reach and effectiveness. Woolaroo is similar from a technical point of view and is even more generalizable to other languages, but it is that same generality that makes it struggle to gain traction. Every language is different, used differently, and taught differently, so a one-size-fits-all solution is often not sufficient.

Conclusion

There will be disagreements within these minority language communities over what data should be allowed to be shared, but researchers should always err on the side of caution and privacy regarding any language data. Language technologies have historically marginalized "treasure" languages, so it is important to direct these technologies toward benefits in reconnecting diasporic endangered language communities and supporting learning efforts if proper communication is maintained with the community about what they consent their data to be used for and what they want from the research.

While still an early and underdeveloped topic of discussion, the application of natural language processing and large language models for language representation raises a multitude of questions about future data consent. Just as prior linguistic data has been collected from speakers who were unaware of how it might be shared online, consent for continued usage of legacy data should not be assumed. Instead, it should be reconfirmed by the origin community for applications like machine learning. Leaders in indigenous data sovereignty like Maggie Walter and Michele Suina provide guidelines and case studies for sustaining indigenous language work. Their methods have inspired NLP researchers to seek a better understanding of production and interaction with content in minority languages and how we might use technology to document languages and build learning platforms. Thus, while I believed that the program I wrote for Kashaya was conducive to bettering our understanding and documentation of the language, I now understand how my positionality should affect my research. Treating language research as a collaborative effort emphasizes where the language data is coming from and prevents it from being reduced to a dictionary and set of grammar rules.

Given how few speakers many of these languages have left, "many tribes, especially smaller ones with fewer resources, rely on non-Native organizations to preserve their languages" (Brewer, 2022). Thus, it is important to enforce responsible research methods and practices to

prioritize the indigenous language communities' usage and incentives. This distinction in the development process explains the difference in popularity of Woolaroo and Kupu. So, by focusing on community driven development, digital tools have even greater potential in revitalizing these endangered languages.

Appendix

This appendix consists of the diagrams, images, and interview questions related to the discussion of using data for minority language revitalization.

Figure 1

Framework for the development of digital language tools for minority and indigenous languages.

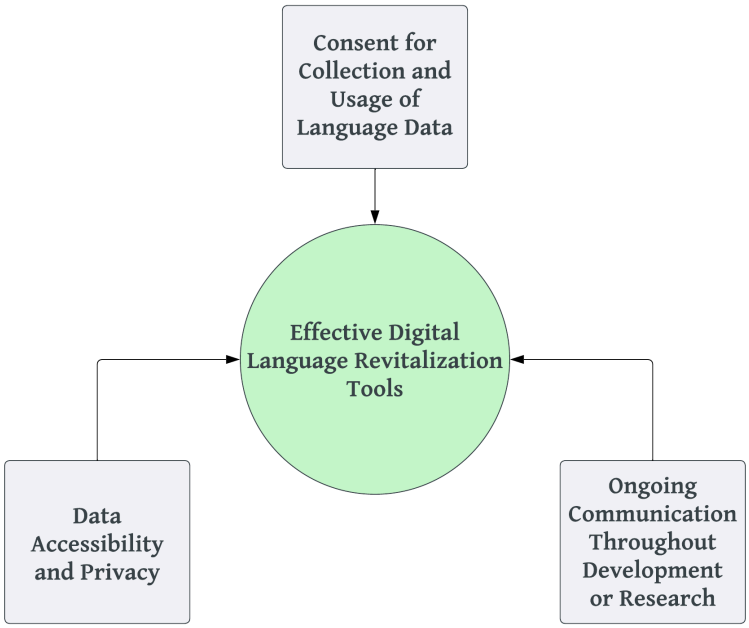
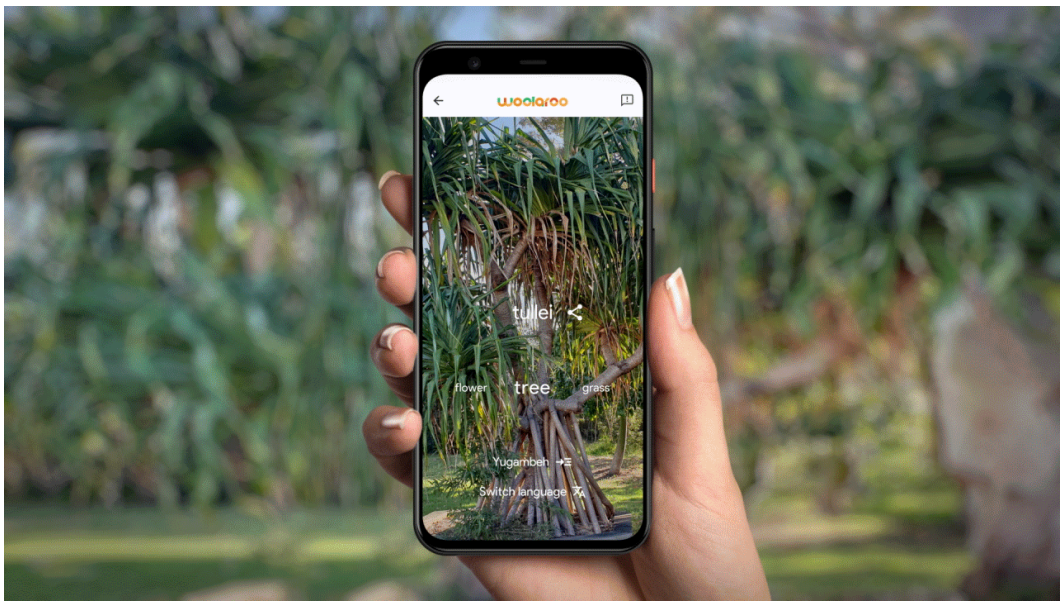


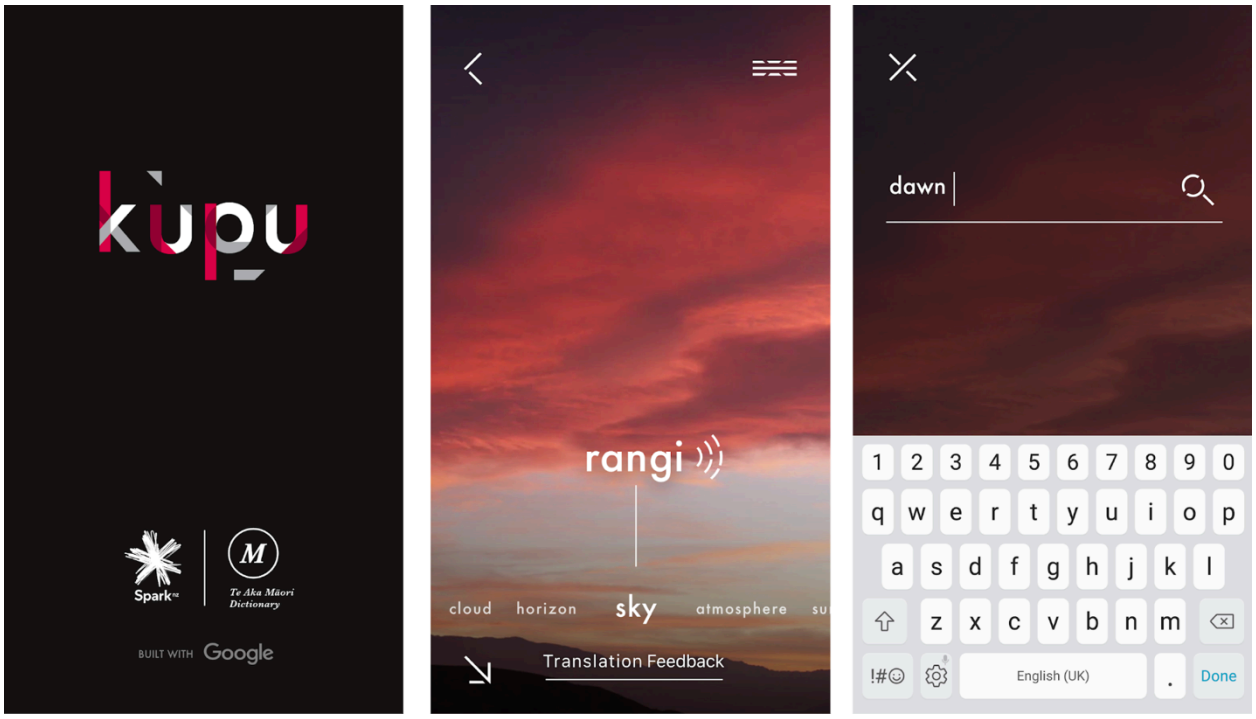
Figure 2

Example usage of the Google Woolaroo visual translation app



Note: Try it here: https://artsexperiments.withgoogle.com/woolaroo/splash

Figure 3
The Kupu visual translation app



Note: Kupu is primarily used in Māori classrooms.

Figure 4
Wesley Leonard Interview Questions

- In an article called *Lingua Nullius: Indigenous Language Learning and Revitalization as Sites for Settler-Colonial Violence*, there is a discussion of how indigenous language learning can be used as fake claims to indigeneity and how it enables settlers to claim that indigeneity. Do you agree with that statement? If so, does that change your thought on online free access to data?
- Do you think it does more harm or good to have this linguistic data publicly available?
- Have there been specific occasions where having this data online has harmed the communities the language belongs to?
- I also read Kornai's paper *Digital Language Death* that talks about how few languages will be able to cross the digital divide. What are your thoughts on his conclusion that 95% of languages can no longer be digital? How do we get these communities to interact more with their data and language online?
- Your paper discusses how it is wrong to store data in third party cloud systems. How else do you propose that they store data that would be better for language documentation?

- If the community itself doesn't have the resources to store it, and it is more accessible to be centralized online, do you think data sovereignty or availability is more important?
- What do you think about language used in machine learning models? For example, with large models created by Google and Open AI for English, this would be trained on existing language content. Do the communities have to further consent to that?
- Does a singular person have permission to give consent to language data which belongs to the community?
- What do you think is the most effective form of language learning and revitalization?

Figure 5
Chance Coughenour Interview Questions

- Can you talk about the idea behind this project and how it first came about?
- What actual language data does Woolaroo store? Just nouns and their dictionary entries?
- How many classes of labels does the vision model have? And if a word in that language doesn't exist, how does an alternate version of the word get created?
- How has the usage been? Are a lot of members from the communities using it?
- How is language support added to the platform and how many languages do you intend to support? Does the model adjust given how thorough a language's dictionary is?
- What are some of the most interesting ways you've seen it be used? For example, have you seen classes or language revitalization programs work with it?
- How do you address data sovereignty and ownership of the language data?
- How involved is the community in the development of a Google Woolaroo language? What dictionaries and data were used to generate the platform? Do you verify if a member of that community is a part of the addition?

References

Bender, E. M. (2011). On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology*, *6*.

Bender, E. M. (2019). The #benderrule: On naming the languages we study and why it matters. The *Gradient*, 14.

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. Retrieved March 12, 2023, from https://www.nbcnews.com/news/us-news/native-american-language-preservation-rcna31396

Cunliffe, D., & Herring, S. C. (2005). Introduction to minority languages, multimedia and the web. *New Review of Hypermedia and Multimedia*, *11*(2), 131–137. https://doi.org/10.1080/13614560512331392186

Davis, J. L. (2017). Resisting rhetorics of language endangerment: Reclamation through Indigenous language survivance . *Language Documentation and Description*, *14*, 37–58.

Frenkel, S., & Thompson, S. A. (2023). 'Not for Machines to Harvest': Data Revolts Break Out Against AI. *International New York Times*, NA-NA.

Google Arts & Culture. (n.d.). *Can you preserve a language?* . Google. Retrieved February 21, 2023, from https://artsandculture.google.com/story/MAWBUhEZEG5pyw

Holton, G., Leonard, W. Y., & Pulsifer, P. L. (2022). Indigenous peoples, Ethics, and Linguistic Data. *The Open Handbook of Linguistic Data Management*, 49–60. https://doi.org/10.7551/mitpress/12200.003.0008

Keall, C. (2019, September 11). *Spark reveals big numbers around its Kupu App*. NZ Herald. Retrieved March 13, 2023, from https://www.nzherald.co.nz/business/spark-reveals-big-numbers-around-its-kupu-app/7MUEMZJ4MFH5M3RNULQF7FAVVM/

Kornai, A. (2013). Digital Language Death. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0077056

Lippi-Green, R. (1990). Language Domination. *English with an Accent Language, Ideology and Discrimination in the United States*, 66–77.

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). https://doi.org/https://doi.org/10.1177/11771801211037672

Noone, G. (2022, April 28). *How New Zealand's Māori people are fighting for their data sovereignty*. Tech Monitor. Retrieved February 2, 2023, from https://techmonitor.ai/policy/privacy-and-data-protection/maori-data-sovereignty-new-zeal and-indigenous

Normanskaja, Y. V., Borisenko, O. D., Beloborodov, I. B., & Avetisyan, A. I. (2022). The software system LINGVODOC and the possibilities it offers for documentation and analysis of Ob-Ugric languages. *Doklady Mathematics*, *105*(3), 187–206. https://doi.org/10.1134/s106456242203005x

O'Connor, R. (2021, May 5). *Woolaroo: A new tool for exploring indigenous languages*. Google. Retrieved February 21, 2023, from https://blog.google/outreach-initiatives/arts-culture/woolaroo-new-tool-exploring-indigenous-languages/

Schulze, I. (2015). Factorizing the sociocultural dimension of language documentation. *Iran and the Caucasus*, *19*(4), 311–333. https://doi.org/10.1163/1573384x-20150403

Shulist, S., & Pedri-Spade, C. (2022). Lingua Nullius: Indigenous language learning and revitalization as sites for settler-colonial violence. *The Canadian Modern Language Review*, *78*(4), 271–288. https://doi.org/10.3138/cmlr-2022-0005

Kupu. Google Play Store. Retrieved from https://play.google.com/store/apps/details?id=nz.co.kupuapp.android&hl=en_US&gl=US.

Petrosyan, A. (2022, August 23). *Internet: Most common languages online 2020*. Statista. Retrieved February 13, 2023, from https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/.

Taiuru, K. (2018, December 18). Why Data is a Taonga A customary Maori Perspective. Retrieved January 21, 2023, from https://www.taiuru.maori.nz/data-is-a-taonga/.

UNESCO. (2010). *Atlas of the world's languages in danger*. United Nations Educational.

Unicode Standard. (n.d.). *Basic questions*. FAQ - Basic Questions. Retrieved February 11, 2023, https://unicode.org/faq/basic_q.html#:~:text=Q%3A%20How%20many%20characters%20are,Unicode%20Standard%20contains%20149%2C186%20characters.

Walter, M., & Suina, M. (2023). Indigenous data, indigenous methodologies and indigenous data sovereignty. In *Educational Research Practice in Southern Contexts* (pp. 207-220). Routledge.