Indigenous Language Technologies & Language Reclamation in Canada

Nathan Thanyehténhas Brinklow[1]**, Patrick Littell**[2]**,**
Delaney Lothian[3]**, Aidan Pine**[2]**, Heather Souter**[4]

Queen's University[1], National Research Council Canada[2], University of Alberta[3], Prairies to Woodlands Indigenous
Language Revitalization Circle[4]
nathan.brinklow@queensu.ca, {patrick.littell, aidan.pine}@nrc-cnrc.gc.ca,
delaney.lothian@gmail.com, hsouter@gmail.com

Abstract

There is an incredible diversity of Indigenous languages spoken and signed in Canada, and Indigenous communities are committed to revitalizing and reclaiming them despite over a century of oppressive governmental policies against their use. Harnessing technology to support these efforts shows promise, but there are common pitfalls to avoid. We briefly describe the social and historical context surrounding Indigenous language technology development and implementation in Canada. We argue that the benefits of technology largely come from those which demonstrate an understanding of the relevant historical context, are tailored to community goals, and emphasize process over product.

Keywords: language revitalization, technology, language reclamation

Résumé

Il existe une grande diversité de langues autochtones parlées et signées au Canada et les communautés autochtones s'investissent dans la revitalization et la reprise de ces langues, malgré des centaines d'années de politiques oppressives contre leur usage. L'utilisation de la technologie est prometteuse, mais il y a aussi des embûches à éviter. Nous décrivons brièvement le contexte social et historique au sujet du développement des technologies langagières autochtones et de leur mise en œuvre au Canada. Nous soutenons que les bienfaits de la technologie se produisent largement lorsque le contexte historique est compris, lorsque les buts communautaires sont intégrés et lorsqu'on met l'accent sur le fait que la démarche est plus importante que le produit.

1. Introduction

Within only a few decades, digital language technology has become a widespread and popular way to help with learning or studying a language. Through greater access to language data and first language speakers, and with high demand from language learners, many strides have been made in creating new technology-based linguistic methods and language learning resources. However, what is far newer is the application of these resources and technologies to the Indigenous[1] languages of people from colonized countries all over the world. A large portion of conventional language resources and technologies are built with the goal of teaching language for reasons of tourism and employment. By contrast, Indigenous language technologies predominantly aim to further efforts of language documentation, revitalization, and reclamation. For this reason, as well as the unique linguistic properties of many Indigenous languages (Littell et. al., 2018), the application of current language technologies and methods to create new ones for Indigenous languages is not a straightforward process. In Canada, there has been a recent increase in funding for the creation of Indigenous language technologies. Though, increased opportunity demands increased critical discussion as to what these technologies can and should look like.

2. Background

Canada is home to approximately 70 unique Indigenous languages belonging to 10 distinct language families. Of these languages, approximately 57% have less than 500 active speakers (Rice, 2019). The decline of Indigenous language transmission in Canada can be attributed to many institutional and societal factors related to colonization, such as residential schools and the Indian Act. Canada's residential school system's explicit purpose was to assimilate Indigenous people into colonial culture by ensuring generations of Indigenous children did not learn their language and culture. These boarding schools were only fully closed "in 1996 after a duration of almost 150 years" (Griffith, 2017). This has profoundly affected the language transmission rates amongst Indigenous peoples in Canada, resulting in fewer and fewer new speakers (Truth and Reconciliation Commission of Canada, 2015).

This history has resulted in Indigenous peoples across Canada developing complex feelings towards learning their ancestral languages and cultures. Through intergenerational trauma from residential schools and other institutional forces, many Indigenous people still feel residual shame around speaking their language (Jenni, Anisman, McIvor, & Jacobs, 2017). Similarly, many

---

[1]Throughout this paper, the term 'Indigenous' will refer to the First Nations, Métis, and Inuit peoples of Canada, and as it pertains to their respective languages and cultures. We follow the Canadian Federal Translation Bureau's capitalization guidelines.

Indigenous people in Canada and around the world do not see learning their ancestral language as being valuable, with majority languages being seen as the language of success and social integration (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003, p. 2). Despite these barriers, there are programs aimed at turning the tide of language loss; from immersion schools and language nests to drop-in classes and mentor-apprentice programs (Dunlop, Gessner, Herbert, & Parker, 2018). The choice Indigenous people all over the country are making to learn and teach their languages, in light of the explicit efforts to suppress them, is a decidedly political act - an act of anticolonial resistance (Pine & Turin, 2017; see also Urla, 2012 & Roche, 2020).

Digital technology is also being employed for Indigenous language revitalization, but its implementation is relatively nascent. Currently there are only a few groups (e.g. the National Research Council of Canada, the University of Alberta, the First Peoples' Cultural Council, the Computer Research Institute of Montréal, etc.) developing technology that focuses on creating building blocks for language work and everyday use, such as fonts and keyboards, spell-checkers, dictionaries, phrasebooks, and predictive text. However, there is relatively little text and speech data available for nearly all Indigenous languages in Canada and so these technologies typically must rely on rule-based approaches (as opposed to data-driven ones).

While there is substantial grammatical variation among Canadian Indigenous languages, they are almost all characterized by a high degree of morphological complexity and polysynthesis (i.e. words are composed of many smaller parts) (Rice, 2008). They are also all substantially different grammatically from the languages in which the foundation of the vast majority of language tools and methods were developed (e.g., English, French, Spanish, etc). This limits the scope of these technologies and leaves a gap between what is possible for Indigenous languages versus what currently exists for more prevalent languages, such as English (Littell et al., 2018).

3. Reframing 'Technology'

In contemporary speech, 'technology' and 'digital technology' are synonymous. In reality, digital technology is built on the shoulders of many other forms of technology. For example, one of the foundational technologies associated with computation is character encoding. Character encoding is how letters and other characters are numerically represented, and is an idea inherited from non-computational technologies like morse code and the telegraph. Given the ill-defined nature of the contemporary use of 'technology', we adopt a more functional definition. We define technology as a *force multiplier*: a tool, idea, object, or technique that allows people to accomplish their goals more quickly or by using fewer resources (see also Rice & Thieberger, 2018). Reframing the definition of technology from something

tool-oriented to something goal-oriented allows technologists and users of technology to be more thoughtful and deliberate about the reasons to employ, or not to employ, a given technology.

When applying a technology to a language revitalization project in Canada, there are many goals to consider. While a primary goal of language revitalization projects is to encourage the use and learning of the language in question, it is rarely the only goal. In an address about the impact of the declaration of the UN International Year of Indigenous Languages, Dr. Lorna Williams stated "in the country that I come from, Indigenous languages were considered, and continue to be considered, of no value [...] our work has been to change this" (Williams, 2019). For many in Canada, like Dr. Williams, the goals of language revitalization transcend linguistic competence; they are connected to identity and community building, cultural resurgence, and broader social goals of self-determination. In recognition of this diversity of goals, some scholars and language advocates are using the term 'language reclamation' in place of revitalization (Leonard, 2012). In other words, the goals of language revitalization are only a subset of the goals of language reclamation.

4. Guiding Principles

In the context of Indigenous language reclamation, we assert that the process of developing language technology is as important as the product. Both the product and the process have the potential to empower or disempower language reclamation communities in equal measure (Alia, 2009, p.173). An example of this danger could be a potential technology that promotes language use but separates a language community from its data; essentially a tool that serves the goal of language revitalization while stifling the goal of language reclamation. This section discusses some of the guiding principles that can help ensure that technological development supports both language revitalization and the social and political goals of language reclamation.

4.1 Technology as a 'MacGuffin'

Technological solutions are frequently oversold in the media (e.g., 'New app is saving endangered language'), and expectations for new technology projects can be unrealistically high. Richard Grounds (2016) echoes this concern, asserting that, "[t]he very notion that these technological solutions somehow represent a kind of comprehensive and easy fix can itself become a problem that stands in the way of finding more effective directions for growing new fluent speakers. And this too often leads to diverting energy away from more effective paths for restoring the strength of our languages." To properly scope a given technology, it is necessary to have a firm understanding that it is *people* that revitalize a language and that technology can merely multiply their efforts. In other words, technology is the icing and not the cake.

It is routine for the 'inherent good' of digital technology to be assumed, regardless of how well it supports language community goals, and for its power and benefits to be exaggerated. In some cases, the technology itself serves as a sort of emblem of prestige much like the partial function of print dictionaries in the Pacific Northwest. While this is a beneficial aspect of technological development (Rice & Thieberger, 2018, p. 236; see also Ogilvie, 2011), new digital technology projects should strive to push past prestige to engage more directly with language communities' goals.

However, even functionally limited technology can concretely further language revitalization goals in its capacity to catalyze and motivate people. As technology development is inherently multi-disciplinary, it can bring together visual artists and craftspeople, musicians, gamers, and others who may not have previously thought of themselves as engaged in language reclamation. This bringing-together of people has sometimes achieved a greater goal than the comparatively minor goal the technology itself was intended to address. This has been seen anecdotally in the development of basic online phrasebooks and dictionaries (see Littell et. al. 2017). In many cases the pedagogical value of these tools for creating new speakers is rather limited; their primary function is as a reference tool. However, when young language learners understand how quickly linguistic data can be published online, it often catalyzes interest in creating more content. The creation of more content for an online dictionary, wordlist or phrasebook provides a focus for these young learners to build relationships and collaborate with elder generations. Strengthening intergenerational relationships is often the fundamental goal of language reclamation projects; a goal which can often surpass the impact of the dictionary or phrasebook itself.

We term such technologies *MacGuffins*: a literary term describing an object of perceived value that moves the plot forward, but which holds little intrinsic value. Many Indigenous language technologies are MacGuffins; building community connections and capacity by their development even if their stated goals are minor or only partially achieved. Put another way, the process of technology development can be valued alongside and even beyond the resulting technology. Realizing this benefit, however, requires planning for it by considering at the outset not just what the product is supposed to do, but who can benefit from their inclusion in the development process.

4.2 Indigenous User Experience

In order for technology to become a supportive stage for Indigenous language reclamation, those involved in the development must have an understanding of how Indigenous language learning differs from conventional language learning and how this will shape the Indigenous user experience. This includes everything from cultural implications of colonization to the appropriate and desired content of a given technology.

There are many ways in which shame may prevent an Indigenous person from learning their ancestral language (Jenni, Anisman, McIvor, & Jacobs, 2017). As mentioned previously, Canada subjected Indigenous people to forced assimilation for generations. It became unsafe and even illegal for some communities to practice their culture and speak their language (Henderson, 2018). Moreover, many people became isolated from their communities through Indian Act control of residency rights which affected, among others, women who married non-Indigenous men and individuals who attended university (Henderson, 2018). Indigenous language learners have cited this residual shame and fear of failure overall as a major obstacle. In this case, technology has a potential role to play in helping learners reach some level of fluency before interacting with speakers (Lothian, Akcayir, & Demmans Epp, 2019).

Another unique factor of Indigenous language technology is that Indigenous languages have historically been strictly oral and this requires consideration and accomodation in the development process. For example, some stories or names are not to be spoken of except by certain people or in certain seasons, and some cultural knowledge is expected to be received in person from elders and knowledge keepers. Accommodating this would require ensuring that the data used to develop content and linguistic models is appropriate and gathered by, or closely with, the relevant community.

4.3 Data Sovereignty & Open Source

Data sovereignty is an area of increasing concern for many Indigenous communities in Canada and internationally (Keegan, 2019). This discussion is rooted in the long history of exploitation by successive colonial governments, which now extends to the potential for exploitation and alienation of Indigenous data (Pool, 2016). In an era of widespread language reclamation activities, many communities identify their language data as a precious resource to be protected. Language technology projects must therefore recognize this reality and ensure that communities are able to protect themselves and their data from colonization in digital spaces (Dyson, Hendricks, & Grant, 2007).

With the staggering level of linguistic diversity in Canada, there is neither the time nor the resources to re-invent technologies each time they are applied to a new language. The language technology ecosystem that supports language revitalization must therefore be open source and well documented. This approach has the advantage of both reducing the level of investment in developing technology, and breaking down some of the barriers for collaboration and participation in the development process. Proprietary solutions, and solutions

which are undocumented to the point of being inaccessible, reinforce the power of a small group of experts at the expense of the larger language revitalization community.

The benefit of open sourcing tools is clear in the low-resource context in which language reclamation is taking place. However, given concerns about data sovereignty as it relates to language reclamation, 'open source' requires a more nuanced application in the Indigenous context, especially at the interface between 'tool' and 'data'. This tension is captured by the Kaitiakitanga License (Te Hiku Media, 2018), which is rooted in Maori community values loosely translated as 'guardianship.' The Kaitiakitanga license model was developed by an Indigenous organization in response to their concern that "by simply open sourcing our data and knowledge, we further allow ourselves to be colonised digitally in the modern world." (Te Hiku Media, 2018). In the Canadian context, the discussion is ongoing and is indicative of the distinction between language revitalization and language reclamation.

5. Conclusion

The prioritization of Indigenous community needs and goals has clear consequences for partners from government, academia, and industry engaged in the development of language technology, including awareness of community cultural values and protocols. The autonomy and agency of Indigenous communities must be recognized in all stages of development as the community decides how to develop and engage with language technologies. This approach allows communities to contribute in meaningful ways to linguistic and cultural continuity through technology as part of language reclamation.

There is clear potential for useful and well-designed Indigenous language technologies in Canada to support language revitalization. However, in the context of language reclamation, responsible technology development must engage with these matters directly. The conversation about Indigenous language technologies is not just about building the right tools, it is about building the tools in the right way—a way that recognizes and affirms the broader social and political goals of language reclamation.

6. Acknowledgements

References

Alia, V., & Berghahn Books. (2009). The new media nation: Indigenous peoples and global communication. New York: Berghahn Books.

Dunlop, B., Gessner, S., Herbert, T., & Parker, A. (2018). Report on the status of B.C. First Nations languages. Retrieved from http://www.fpcc.ca/files/PDF/2010-report-on-the-status-of-bc-first-nations-languages.pdf

Dyson, L. E., Hendriks, M., & Grant, S. (Eds.). (2007). Information technology and Indigenous peoples. Hershey, PA: Information Science Publishing.

Grounds, Richard A. 2016. Indigenous Perspectives and Language Habitats. Paper presented to the *International Expert Group Meeting on Indigenous Languages: Preservation and Revitalization*. United Nations, Department of Economic and Social Affairs, New York, January 19–21. http://www.un.org/esa/socdev/unpfii/documents/2016/egm/ Paper_Grounds2.pdf.

Griffith, J. (2017) Of linguicide and resistance: Children and English instruction in nineteenth-century Indian boarding schools in Canada. *Paedagogica Historica*, 53:6, 763-782, DOI: 10.1080/00309230.2017.1293700

Henderson, W. (2018). Indian Act In *The Canadian Encyclopedia*. Retrieved from https://www.thecanadianencyclopedia.ca/en/article/indian-act

Jenni, B., Anisman, A., McIvor, O., & Jacobs, P. (2017). An exploration of the effects of mentor-apprentice programs on mentors' and apprentices' wellbeing. *International Journal of Indigenous Health*, 12(2), 25. https://doi.org/10.18357/ijih122201717783

Keegan, Te Taka. (2019, June 25). Issues with Māori sovereignty over Māori language data [Video File]. Retrieved from http://video.web.gov.bc.ca/public/fpcc/letlanguageslive.html

Leonard, W. Y. (2012). Framing language reclamation programmes for everybody's empowerment. *Gender and Language,* 6(2), 339–367.

Littell, P., Kazantseva, A., Kuhn, R., Pine, A., Arppe, A., Cox, C., & Junker, M. (2018). Indigenous language technologies in Canada: Assessment, challenges, and successes. In J. L. Klavans (Ed.), *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2620–2632). Retrieved from https://www.aclweb.org/anthology/W18-1921

Littell, P., Pine, A., & Davis, H. (2017). Waldayu and Waldayu Mobile: Modern digital dictionary interfaces for endangered languages. In Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages (pp. 141-150).

Lothian, D., Akcayir, G., & Demmans Epp, C. (2019). Accommodating Indigenous People When Using Technology to Learn Their Ancestral Language. *International Workshop on Supporting Lifelong Learning at the 20th International Conference on Artificial Intelligence in Education (AIED)* (Vol. 2395

pp. 16-22), Chicago, Illinois, USA. CEUR-Workshop Proceedings.

McIvor, O., & Anisman, A. (2018). Keeping our languages alive: Strategies for Indigenous language revitalization and maintenance. In Y. Watanabe (Ed.), *Handbook of Cultural Security* (pp. 90–109). https://doi.org/10.4337/9781786437747.00011

Ogilvie, S. (2011). Linguistics, lexicography, and the revitalization of endangered languages. *International Journal of Lexicography*, 24(4), 389-404. doi:10.1093/ijl/ecr019

Pine, A., & Turin, M. (2017). Language Revitalization. Oxford Research Encyclopedia of Linguistics. Retrieved from, https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-8

Pool, I. (2016). Colonialism's and postcolonialism's fellow traveller: The collection, use and misuse of data on Indigenous people. In T. Kukutai & J. Taylor (Eds.), *Indigenous Data Sovereignty* (pp. 57–76). Retrieved from https://www.jstor.org/stable/j.ctt1q1crgf.11

Rice, K. (2019). Indigenous languages in Canada. In *The Canadian Encyclopedia*. Available at https://www.thecanadianencyclopedia.ca/en/article/aboriginal-people-languages

Rice, K., & Thieberger, N. (2018). Tools and technology for language documentation and revitalization. In K. L. Rehg & L. Campbell (Eds.), *The Oxford Handbook of Endangered Languages.* https://doi.org/10.1093/oxfordhb/9780190610029.013.13

Roche, G. (2020). Language revitalization and radical politics. Language on the Move. https://www.languageonthemove.com/language-revitalization-and-radical-politics/

Te Hiku Media. (2018). Kaitiakitanga License. Retrieved from https://github.com/TeHikuMedia/Kaitiakitanga-License/blob/master/LICENSE.md

Truth and Reconciliation Commission of Canada. (2015). Honouring the truth, reconciling for the future. Retrieved from http://www.trc.ca/assets/pdf/Honouring_the_Truth_Reconciling_for_the_Future_July_23_2015.pdf

UNESCO Ad Hoc Expert Group on Endangered Languages. (2003). Language vitality and endangerment. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf

Urla, J. (2012). Reclaiming Basque: Language, nation, and cultural activism. University of Nevada Press.

Williams, L. (2019, Dec 6). "Technology: Is it a help or a hindrance?" [Video File]. Retrieved from http://webcast.unesco.org/live/vod/2019/ci/20190612_ci_room-02/en/