

Decolonising Speech and Language Technology

Steven Bird
Northern Institute
Charles Darwin University

Abstract

After generations of exploitation, Indigenous people often respond negatively to the idea that their languages are data ready for the taking. By treating Indigenous knowledge as a commodity, speech and language technologists risk disenfranchising local knowledge authorities, reenacting the causes of language endangerment. Scholars in related fields have responded to calls for decolonisation, and we in the speech and language technology community need to follow suit, and explore what this means for our practices that involve Indigenous languages and the communities who own them. This paper reviews colonising discourses in speech and language technology, and suggests new ways of working with Indigenous communities, and seeks to open a discussion of a postcolonial approach to computational methods for supporting language vitality.

1 Introduction

So much of the templates, the guidelines, the university courses, have been created along the lines of a colonising language. Our task is to figure out how to be able to speak our languages, how to teach our languages, how to maintain our languages, how to normalise our languages, from our ways, from our perspective, from our point of view. – Lil’watul Lorna Williams, Lil’wat First Nation

This is how Lorna Williams addressed 300 people at *Language Technologies for All* (Williams, 2019). We had been summoned to the UNESCO headquarters in Paris with these words:

Everyone should have the possibility to get access to Language Technologies in his/her native languages, including Indigenous languages. The challenge is twofold: (i) preserving culture through language while (ii) allowing communication across languages. Languages that miss the opportunity to adopt Language Technologies will be less and less used, while languages that benefit from cross-lingual technologies such as Machine Translation will be more and more used (European Language Resources Association, 2019, capitalisation as in original).

The world’s 370 million Indigenous people (United Nations, 2010) have been passing down their knowledge over countless generations, commanding multiple languages in the process. But now, Indigenous people are apparently unable to go on without speech and language technologies.

By the time she addressed the conference, Williams had witnessed two days of “tech-evangelism” (Birhane, 2020, 394). Yet she did not call on her community to get with the program. She said: *the task is to figure out ... how to maintain our languages from our ways, our perspective, our point of view.*

Two years earlier, in Australia, the Federal Minister for Communication and the Arts announced new funding for Indigenous languages. He brought together the nation’s linguists and language workers with the goal of “facilitating ideas and skills to *advance digital solutions* to protect, preserve and celebrate Indigenous languages” (bit.ly/34LFfkk). In a session titled “Digital Approaches that Work” a 23-year-old Anaiwan man, Callum Clayton-Dixon, spoke out: “we need to prioritise ‘on land’ over ‘online’... you don’t revive a language with an app but with people reconnecting with people and country.” From the applause and the endorsements, his was a widely held view.

This work is licensed under a Creative Commons Attribution 4.0 International Licence.
Licence details: <http://creativecommons.org/licenses/by/4.0/>.

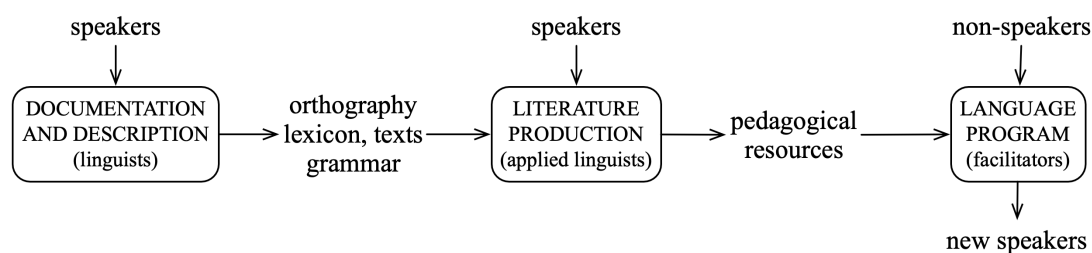


Figure 1: The Language Documentation Mechanism: linguists (present and past) create language resources that support the construction of pedagogical resources to be used in language revitalisation

For 25 years, many linguists have presented “indigenous language endangerment as a crisis that requires intervention” (Nevins, 2013, 21). Like many, I have taken this crisis as an golden opportunity for technology, prioritising data capture over local self-determination, in such areas as digital preservation (Bird and Simons, 2003), capacity building (Bird, 2010; Bird and Chiang, 2012; Bird et al., 2013), and app development (Bird et al., 2014a; Bird et al., 2014b).

Five years ago I entered the Indigenous country of Arnhem Land in northern Australia imagining language capture on an even larger scale. In the course of living and working in an Aboriginal community, I began to participate in the language activities of a school, a ranger program, and an arts centre. I came to see how commodifying Indigenous languages as data alienates speakers (Dobrin et al., 2009), and how technology does not address the social injustices that underly language endangerment (Srinivasan, 2017). I saw how a preoccupation with data and technology might reenact the causes of language endangerment.

This is a paper about colonising practices in speech and language technology, ways to identify them, and to shift to decolonising practices. I begin with illustrations and analysis of the ways that technology may be thought to save languages (Sec. 2), before reviewing some approaches to decolonising language work (Sec. 3). Finally, I present a new framing for speech and technology work in terms of participatory design, and a variety of concrete actions for consideration (Sec. 4).

I acknowledge my positionality as a western researcher of settler-coloniser heritage. I am committed to an ongoing process of decolonising my solidarity with the community. While I have consulted my Indigenous advisors on many of the topics discussed here, I do not speak for them. I acknowledge the discomfort (and irony) of the othering that is inherent in my use of terms like “they” and “them,” and the fact that I did not manage to find a way of coauthoring such a paper with an Indigenous colleague.

2 Technology Saving Languages

Since the declaration of the “world’s languages in crisis” (Hale et al., 1992), many linguists have taken on the task of documenting endangered languages (Himmelmann, 1998), with particular focus on so-called “Indigenous languages” (Walsh, 2005; Merlan, 2009). Documentation is believed to play into revitalisation (Himmelmann, 2006). The mechanism has not been made explicit, but perhaps Figure 1 comes close. Language programs are often cast as an antidote to endangerment, when both are “facets of the same process of political integration” (Nevins, 2013, 18ff). We see this in Figure 1 in the absence of agency of speakers, and in the construct of “pedagogical resources” to be deployed in formal programs.

There is scant evidence after 25 years that documentation has played a significant role in saving languages (Nathan and Fang, 2014). Indeed, UNESCO identifies many factors that impact the survival prospects for a language, such as intergenerational transmission (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003). Nevertheless, speech and language technologists have bought into the documentary agenda, e.g., “Systems constructed with zero expert resources [can] help field linguists document endangered languages, by providing tools to semi-automatically analyze and annotate audio recordings using automatically discovered linguistic units” (Dunbar et al., 2017). This cooperation is hardly surprising when the documentary agenda aligns so neatly with the data-centric approach in speech and language technology. We can harness linguists and speakers to create training data, and arrive at a technological version of the mechanism (Fig. 2). There are at least three problems with this approach.

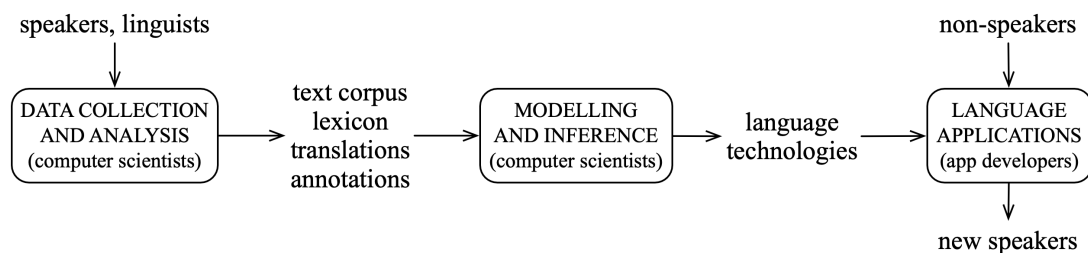


Figure 2: The Language Technology Mechanism: how technologies might enable language revitalisation

2.1 “Zero Resource Scenario”

“Colonization is ... on the one hand ‘experimental civilization,’ where European ideas could be tested with new data, and, on the other hand, ‘applied civilization,’ where European ideology could be used for ‘the good of mankind” (Samarin, 1984, 436). “Discovering” is tantamount to “saving” (Nevins, 2013, 32). “A love of exploration is part of the motivation of many ‘diversity linguists’: newly described languages constitute unknown land to explore” (Michaud et al., 2018, 20). In a neocolonial moment, automatic methods set sail to discover “previously unseen languages” (Vetter et al., 2016).

A speech scientist contacted a linguist for his data: “the scenario was that nothing was known about the language, and the data set available consisted solely of audio recordings of sentences plus translations into other languages” (Michaud et al., 2018, 8). The scientist did not need the linguist’s transcriptions, dictionary, phoneme inventory, or knowledge. Nor did he need to engage with speakers of the language. A fully-automatic, universal solution based on sanitised data was preferable to the alternative of working with people who have knowledge about the language and concerns for its future.

This “zero resource scenario” disenfranchises local knowledge authorities (Verran and Christie, 2007, 215). It is a version of the cliché of “new technologies saving ancient languages,” perpetuating colonial dichotomies of advanced vs. primitive, of domesticated vs. wild (Goody, 1977; Nevins, 2013).

When we are unable to hear what linguists and speakers would tell us, we must fall back on unsupervised methods, like discovering a phone inventory from speech alone (Kempton and Moore, 2014; Vetter et al., 2016; Adda et al., 2016; Müller et al., 2017), or segmenting phone sequences into “words” (Johnson and Goldwater, 2009; Elsnér et al., 2013). Yet after centuries of colonisation, missionary endeavours, and linguistic fieldwork, all languages have been identified and classified. There is always a wordlist. We know the language family. Related languages have been studied. There are texts and translations. There may be linguists and speakers. In short, we do not need to “discover” the language *ex nihilo* (L1 acquisition) but to leverage the available resources (L2 acquisition). Indigenous languages do not deserve this zero resource framing.

2.2 “Gold Data”

From this position of language as lexico-grammatical code divorced from social functions, practitioners shift into extractive mode. Systems need refined data, i.e., monologic speech recordings made in a clean acoustic environment (Hermes and Engman, 2017) with none of hallmarks of embodied communication (Tedlock, 1983). The messy and incomplete processes that created the data are hidden from view (Kleinman and Copp, 1993; Star, 2007; Himmelmann, 2018). Transcriptions are seen as neutral, even though transcribing is theory-laden observation (Ochs, 1979; Cucchiari, 1993; Bucholtz, 2007; Hermes and Engman, 2017). Quantity and quality are reduced to scalar measures: data size and error rate.

The preoccupation with getting all speech into textual form (Mosel, 2006; King, 2015) and delivering “exhaustive transcriptions that are faithful to the audio” (Michaud et al., 2018, 12) misconstrues the nature of transcription and creates a monumental and menial workload for linguists and speakers. By hypervalorising faithfulness, it causes speakers the discomfort of seeing their disfluencies and mistakes enshrined in frozen documents (Marten and Petzell, 2016), when they prefer to make exemplary artefacts (Wilkins, 2000). The destination – the archive – is itself a legacy of the colonial program to institutionalise knowledge (Dourish and Mainwaring, 2012, 133).

2.3 “Unwritten Languages”

In the discourse of ‘endangered languages’, it turns out that writing – its presence or its absence – is what makes all the difference... this discourse about language(s) establishes a horizon, a line, on the far side of which is oblivion... The discourse of language endangerment situates the act of writing precisely astraddle that line, endowing writing with the power to move (a) language, word by painstakingly transcribed word, from one side of that metaphysical divide to the other (Moore, 2006, 313).

To frame an oral vernacular as an “unwritten language” is to set an agenda: it needs to have text-based resources and applications (Arppe et al., 2016; Maheshwari et al., 2018). The absence of the technology of writing aligns with other negatives: to be *under-resourced*, a language *lacks* what is required for creating speech and language technologies (Krauwert, 2003). In this one-dimensional view, Indigenous and endangered languages are *acutely under-resourced* (Jimerson and Prud’hommeaux, 2018).

With the best of intentions we can imagine “developing speech recognition (and other technologies like machine translation) systems for literally all languages in the world” (Besacier et al., 2014, 86). This automatic translation connects with another colonising frame: universal access to knowledge (Dourish and Mainwaring, 2012). In short, the view seems to be that major western languages are standard-bearers, and that Indigenous languages need the standard technologies. Moore’s analysis of the hegemonic status of writing applies equally well to speech and language technology.

3 Decolonising Language Work

Linguists have been deepening their engagement with speech communities (Hinton and Hale, 2001; Rice, 2009; Sapién, 2018), calling out the emotive and moralistic tropes used to legitimise the documentary approach (Dobrin et al., 2009; Perley, 2012), criticising nostalgia and sentimentalism (De Swaan, 2004; Good, 2018), observing the professional narrowness of the focus on linguistic structures (Dorian, 2002, 136), and bemoaning the prevalence of linguistic material that is not accessible to speakers (Gerdt, 2010). This connects to a broader program of decolonising research (Dourish and Mainwaring, 2012; Nevins, 2013; Christie, 2014; Land, 2015; Stebbins et al., 2017), and building social justice and self-determination into the research agenda (Smith, 1999, 204).

A starting point is the recognition that we operate in a postcolonial situation, where language loss is often the byproduct of oppression (Nevins, 2013, 25), particularly for Indigenous people in liberal democratic settler states (Merlan, 2009). In Australia, “government policies and practices denigrated Aboriginal languages, and children were punished for speaking Aboriginal languages” (Amery, 2007, 328). Governments well understood the function of Indigenous languages in sustaining identity and connection to the country, which is why “languages are so often the target of oppression in processes of colonisation or displacement” (Stebbins et al., 2017, 39). Language activities have the potential to support resilience and wellbeing, and elevate local voices (Bradley, 2010; Whalen et al., 2016; Stebbins et al., 2017, 41f). However, just as often these activities respond to externally perceived problems and deliver “solutions” that create new problems for local people (Moran, 2016; Stebbins et al., 2017, 21).

In contrast, decolonising practices create spaces where “the oppressed [can] begin to believe in themselves” (Freire, 1970, 39). They dislodge white privilege to access knowledge (Stebbins et al., 2017, 39). They acknowledge that: “Indigenous peoples have the right to self-determination. By virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development” (Article 3, UN Declaration on the Rights of Indigenous Peoples).

What does self-determination look like in the case of language work? “To imagine self-determination [then, is] to imagine a world in which Indigenous peoples become active participants” (Smith, 1999, 127). As a result, “the community must be the driver and driving interest of the research and language projects” (Stebbins et al., 2017, 52). Indigenous people have the option to abandon their languages (Kulick, 1997; De Swaan, 2004). But if they choose to develop their languages, they have a right to self-determination, engaging outside experts “to support a program or strategy the community has already settled on rather than to develop a discrete project based on [the outsider’s] expertise” (Bedford

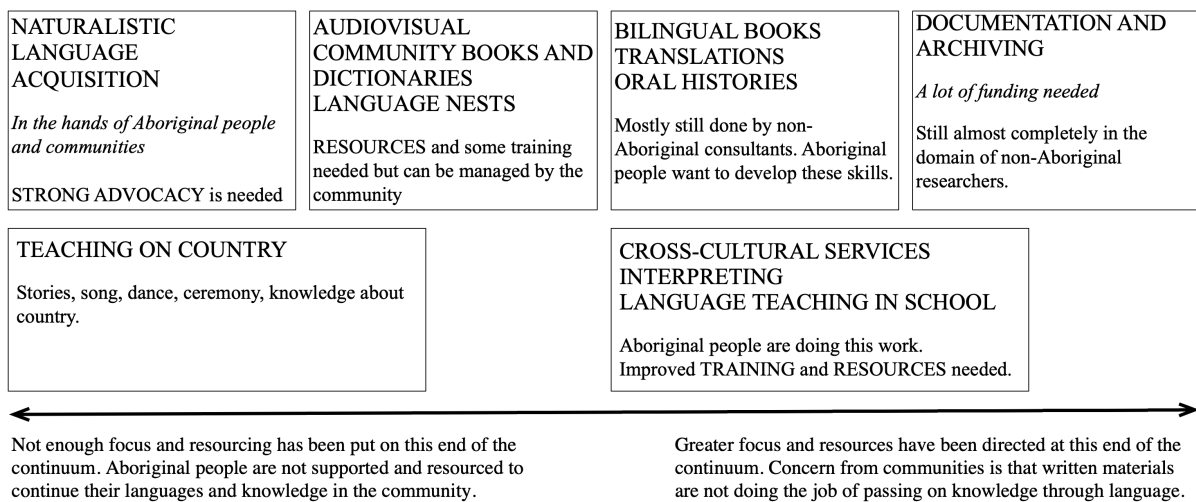


Figure 3: Language Continuation Continuum: communities want support for activities towards the left, while national funding programs and linguists favour activities on the right (Bedford and Casson, 2010)

and Casson, 2010, 80). This latter group analysed activities on an axis of self-determination (Fig. 3). We can see that self-determination hits up against the problem of accountability for funding (Stebbins et al., 2017, 245), and cherished beliefs about pedagogy (Nevins, 2013, 58ff).

The target of revitalisation is also contested. In reality, the revitalised language may differ substantially from the ancestral one (Zuckermann, 2020). Emerging codes are pressed into service: “while Kriol speakers have massively replaced Australian lexical forms by English-like forms, they have been able to do so without modifying the meanings expressed in Australian languages to the same extent” (Ponsonnet, 2019, 40). Salient aspects of the ancestral life-world persist.

Furthermore, many Indigenous communities are *diglossic*, using vernacular languages for transmitting local knowledge, and vehicular languages for participating in the national economy and accessing western education and health systems. Recent thinking on sustaining language use encourages us to address not the language community but the *speech* community, recognising its “multiple communicative regimes” (Nevins, 2013, 28), its “repertoire of languages” (Lewis and Simons, 2016, 35ff). Code-switching may not be a threat, but it may support the learning of ancestral languages, via the practice of “pedagogical translanguaging” (Cenoz and Gorter, 2017; Seals and Olsen-Reeder, 2020).

In spite of these nuanced understandings of the target, many external scholars advocate for the pure ancestral language, even when this is counterproductive (Dorian, 1994). Here is a place where linguists may have a blind spot, criticising the prescriptivism of laypeople while indulging in it themselves (Stebbins et al., 2017, 44f). “The discourse of hyperbolic valorization converts endangered languages into objects more suitable for preservation in museums patronized by exceptionally discerning elites than for ordinary use in everyday life by imperfect human beings” (Hill, 2002, 120).

In places where people wish to connect with their ancestral language, several programs have been devised, such as the Master Apprentice Program (Hinton, 2001; Grenoble and Whaley, 2006, 50ff). The goal may not be fluency: “a belief [in] a goal where everyone in the community will speak Karok to each other again is unrealistic. Language revitalisation in areas of language diversity and small populations is going to be very different than for languages like Hebrew, Hawaiian and Maori” (Hinton, 2015, 39.00m). An Australian survey found that most people were less interested in proficiency, and understood that the goal of language activities was “to strengthen people’s connection with their language and culture, to build a sense of identity and wellbeing, and to increase language awareness” (Marmion et al., 2014, 19).

To be sure, the speech and language technology community brings a useful toolbox (Stebbins et al., 2017, 48). But it also brings unhelpful assumptions about goals and methods, and a habit of algorithmic and technological colonisation (Grossman, 2014; Mohamed et al., 2020). How do we decolonise our approach?

4 Codesigning Speech and Language Technologies with Communities

4.1 A Cross-Cultural Encounter

“Ethical research must begin by replacing Eurocentric prejudice with new premises that value diversity over universality” (Battiste and Henderson, 2000, 133). There is no way to avoid a cross-cultural encounter (Nevins, 2013, 36ff), where we build relationships that are marked by equality and reciprocity (Brereton et al., 2014; NHMRC, 2018). We locate and grow a “third space,” a culturally safe place where discomfort can be managed and leads to productive tension and hybrid outcomes (Somerville and Perkins, 2003; Christie and Verran, 2014). This may function as a “recognition space” where Indigenous and external actors can interpret each other’s agendas and identify shared tasks that serve both (Fig. 4).

In Australian Aboriginal communities this often begins with sitting with people, having a yarn over a cup of tea by a fire. There is no interrogation, just sharing stories about families, the place, and what is happening with the language. We could talk about what it might mean to sustain or forget this language (Nevins, 2013, 8). The newcomer might share her own family history of intergenerational disruption of language transmission, and what this has meant. Such “yarning” is a research method in itself (Bessarab and Ng’andu, 2010; Ober, 2017).

In this context, we rethink the goals and practices of technology development (Dourish and Mainwaring, 2012, 139ff). We may use technological artefacts to prompt design conversations (Verran et al., 2007; Soro et al., 2016). We can understand such engagements “less as ways that designers can formulate needs and measure outcomes, and more as shaping and staging encounters” (Irani et al., 2010, 1317), an iterative process (Dourish et al., 2020). When it comes to capturing language materials, we must take seriously the sovereignty of local people over their data (Bow, 2019; Walter and Suina, 2019). In sum, we enter a research process that supports recovery from the colonial legacy (Stebbins et al., 2017, 47ff), beginning with self-determination (Battiste and Henderson, 2000, 132ff).

4.2 Identifying community goals

It is no small thing for an outside “expert” to enter an Indigenous community and become the latest in a series of external engagements (Dorian, 2002; Nevins, 2013; Grinevald, 2007, 42f). Yarning may lead on to semi-structured interviews and coding (Charmaz, 2014), where we come to understand, from the “ground up,” how languages and language revitalisation activities function locally, and identify suitable indicators of progress. This process may challenge the language ideology of the outsider, of language-as-data, language-as-tool, language-as-lexicogrammatical-code, or even language-as-emblem-of-identity. For example, outsiders may be challenged to broaden the scope of the engagement from capturing the lexicon and grammar of an ancestral language to studying and supporting “language in use” (Stebbins et al., 2017, 21), or communicative competence across a repertoire of languages (Nevins, 2013, 59).

It is possible that the parties in a collaboration may view its artefacts differently. A feature of decolonising approaches is to embrace multiple perspectives (Dourish and Mainwaring, 2012, 140). In particular, we can understand speech and language technologies as “boundary objects” (Star and Griesemer, 1989; Star, 2010). For example, a dictionary app may represent an access point for one participant, and a marker of recognition in the digital realm to another. And this latter is no less valuable, given that prestige is a factor in language shift (Guérin, 2008; Pine and Turin, 2017).

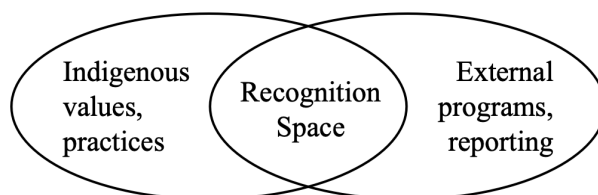


Figure 4: The Recognition Space, where Indigenous values and practices can be made visible to external agencies, e.g., in the form of social indicators, after (Taylor, 2008, 116)

Embracing community goals may entail a shift away from the customary agendas of linguistics and natural language processing, to broader understandings of the potential for computational methods. This is already underway in established workshop series including ComputEL, SLTU, CCURL, LaTeCH. To the extent that speech and language technologies count as instances of ubiquitous computing technologies, we may follow the recommendations of Dourish and Mainwaring (2012).

4.3 The Sustainable Use Model

In their Sustainable Use Model, Lewis and Simons (2016) set out four levels of sustainability. Their work began with Fishman's 8-level GIDS scale, the most widely known measure of language endangerment. Lewis and Simons (2010) added more levels to create the 13-level Expanded Graded Intergenerational Disruption Scale (EGIDS), encompassing all languages and better serving language revitalisation efforts. Of these, there are just four sustainable levels: *Sustainable Literacy* (EGIDS 4) "The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education." *Sustainable Orality* (EGIDS 6a) "The language is used for face-to-face communication by all generations and the situation is sustainable." *Sustainable Identity* (EGIDS 9) "The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency." *Sustainable History* (EGIDS 10) "The language has been documented adequately and the documentary materials are safely and reliably archived for future access."

Lewis and Simons analyse sustainability at any of these levels under the "FAMED conditions": (a) there are *functions* for the language, also known as domains of use; (b) there are means for *acquiring* the language at the level required for performing the functions; (c) there is *motivation* for acquisition, the perceived benefits of learning the language; (d) there is a favourable policy *environment* for the language created by government and non-governmental organisations; and (e) there are norms of language use that *differentiate* uses of the threatened language from uses of locally more dominant languages.

We can apply these conditions to understand that the scenario in Figure 1 is unlikely to succeed by itself. While it covers acquisition and environment, it does not establish functions, motivation, or differentiation. There are many more ways in which technologies designed for sustaining language use may fail to deliver on their promise: focussing on a language in isolation from the linguistic repertoire of a community; emphasising the language as a lexico-grammatical code to be mastered over and above its function for transmitting life-crucial knowledge; and investing in an overly ambitious sustainability level, wasting resources and generating a sense of inadequacy. Lewis and Simons' remedy is for all language work to privilege community aspirations, realistic goals, and steady shift up the EGIDS scale.

4.4 Enlarging our understanding of evaluation

Speech and language technologists have bought into a theory of language continuation (Fig. 2), while only building prototypes that address the initial stages, and only performing the most simple-minded kinds of evaluation. Where are the full implementations, deployed in robust software, in active use to capture primary data, leading to curated language products that are being mobilised in speech communities? Might technology be implicated in other processes like intergenerational transmission, addressing threats to Indigenous languages other than the lack of documentation (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003)?

Evaluation is broader than reductionist accounts of the performance of system components. For example, evaluation is present at the start of any project, when it is considered for feasibility, for funding, and for ethics approval. Evaluation is also a device for engagement, because it leads us to ask how speech and language technologies relate to "knowledge sharing, joint engagement in language maintenance activities, or other kinds of interactionally-defined achievements" (Dobrin et al., 2009, 43).

Privileging local agency. Looking at the headlines and stories that come out of the work, we ask: to what extent do they elevate local voices? The headline might be: "One of the world's 7,000 languages dies every three months: Can apps help save them?" (Steinmetz, 2018). Imagine a different framing: "Indigenous people revitalise their language, leveraging technology." Here is a place to evaluate whole projects: how much funding goes to Indigenous participants, which products are truly delivered to community members, how well can these products be locally sustained and extended after the project

finishes, how do the participants themselves evaluate the project, who are the media stories about? Finally, to the extent that participants have articulated a vision for a desired future, we can inquire about the contribution of our technological artefacts towards realising that future (Taylor et al., 2016).

Identifying differentiated functions. We return to the FAMED conditions (Lewis and Simons, 2016). Technology opens up new domains of use (i.e. functions), it may support learning (i.e. acquisition), and it may provide a hook (i.e. motivation). The existence of funding suggests there is a favourable policy context (i.e. environment). The last condition, *differentiation*, may be the most challenging. Technology deployments must support differentiated functions, not simply replicate functions that are already established in the locally dominant language. Otherwise, it is too easy for participants to slip back into the dominant language.

Finding the human in the loop. Most empirical work in speech and language technology is built on an input-output conception of a linguistic task. Human involvement ceases upstream, when the corpus was created. Systems are evaluated according to their ability to replicate human judgements. A “shared task” does not refer to collaboration with speakers but competition with like-minded researchers. Could we evaluate systems on their ability to support culturally meaningful collaborative work (Verran et al., 2007), i.e., actual shared tasks? Elsewhere, I have tried to recast transcription in this way (Bird, 2020).

Expanding people’s capabilities. When it comes to language, many Indigenous people value the connections, identity, and wellbeing that language enables, over and above individual proficiency (Sec. 3). “The pathways differ from group to group and time to time, depending on the situation... revitalisation has no stopping point where someone can say we have succeeded” (Hinton, 2015). So, following Sen (1999) we can ask: how do speech and language technologies serve to expand people’s capability to lead the kind of lives they have reason to value?

Identify underlying mechanisms. A way to approach this “how” question is offered by the framework of Realist Evaluation (Pawson, 2013). We use abductive reasoning to identify underlying *mechanisms* that explain how and why something produces (or fails to produce) a desired *outcome* in a given *context*. Mechanisms can be understood as the interplay between the resources which a program makes available and how participants respond. Context refers to the factors that influence whether or not a mechanism works. In any technology deployment there is social contingency: it is not so much the technology but people’s response to it that is responsible for change (Wright and McCarthy, 2004, 30).

The main activities in a realist inquiry are developing a theory that explains how something works, and testing the theory through empirical research. The product of a realist inquiry is a better theory, that is, a more nuanced understanding of how outcomes are caused, what causes them, and who tends to achieve which outcomes, and why. This kind of knowledge is more useful than a simple pass/fail judgement, because it provides new knowledge which is relevant to the context and experiences of participants. The methods may be qualitative or quantitative (Pawson, 2006, 18). Initial theories may come from the literature, or from participatory design processes.

4.5 Getting started

Build relationships. The parameters of ethical engagement include respect, reciprocity, and equity. It is difficult to imagine how one could satisfy such requirements without first establishing relationships with local people. And this cannot be superficial: Indigenous people are already sensitised to extractive engagements, and to researchers who leverage them for virtue signalling or for building their careers. When this fundamental shift is made, it is noticed. It opens doors. More than this, relationships may become integral, to the point where we are designing “relational language technologies” (Taylor et al., 2019).

Become sensitised to local agency. Much of the rhetoric surrounding computational approaches to endangered languages turns on the power of technology to save languages. This is not just a cliché and false advertising, it fails to recognise people’s agency in speaking whatever language(s) they choose. Reversing language shift is difficult (Fishman, 2001). It ultimately requires community participation, so why not start there? What are people already motivated to do, and how might speech and language technology amplify this?

Target vehicular languages. Vehicular languages are the codes people use to communicate outside their linguistic community. Speakers already use such languages for conducting the functions typically envisaged for speech and language technologies. A high-impact starting point would be to support vehicular and provincial languages (EGIDS 3, 2), and to improve the robustness of technologies for local varieties of national and international languages (EGIDS 1, 0).

Prioritise knowledge transmission. Indigenous communities may be more concerned about the loss of knowledge than the loss of language per se (Lewis and Simons, 2016, 35ff). Documenting and transmitting life-crucial knowledge can be viewed as an instance of *computer supported cooperative work* (Grudin, 1994; Hanke, 2017; Bettinson and Bird, 2018).

Mobilise the archive. Preoccupation with data collection may neglect materials that have already been collected. It may be possible for these to be disseminated, and such dissemination may leverage a popular use of mobile devices in some Indigenous communities, namely, accessing and sharing digital media (Auld et al., 2012; Carew et al., 2015). This demonstrates the value of digital preservation to people, and may stimulate discussion, as when people want to talk about a person or activity that is captured in a photo or recording.

Support oral language learning. In the realm of speech and language technology, the focus of learning has been the machine, in order to use human labour efficiently. In the realm of language learning technologies, the focus has often been on overcoming the low-resource situation to create pedagogical content (Ward and Genabith, 2003; Ward, 2018), resulting in text-based solutions (Hermes and King, 2013; Katinskaia et al., 2017). There is an opportunity to support learners in creating their own oral language materials on the fly and leveraging such materials in their learning.

4.6 Not giving up

It would be a counterproductive if members of the speech and language technology community decided to drop their interest in Indigenous languages, because of the inherent difficulties touched on here. To be sure, the risks are significant. No community speaks with a single voice, and in building relationships we can unwittingly align ourselves with agendas, clans, and gate-keepers. We may invest in building relationships over years, only for our cherished project to be shut down (Nevins, 2013, 12ff). Anything that touches on how a language is written can be particularly sensitive (Bird, 2001). However, there are synergies. The speech and language technology community is accustomed to certain ways of working that may align with work in Indigenous contexts.

First, as recipients of government funding, we are accustomed to writing aspirational research proposals that carve out a space where future, unpredictable developments can occur. The same can be true when seeking funding for Indigenous engagements; the project can be elaborated with Indigenous participation, and we can view the proposal document as “white man’s paperwork” which simply creates a space where the real, collaborative work can happen.

Second, members of the speech and language technology community routinely construct a sequence of funded projects, each next project building on the last and heading in new directions. Similarly, one can envisage a series of projects, beginning with a small pilot which prioritises relationship building, leading to a series of collaborative proposals where Indigenous participation is continually expanded. We may support local efforts to obtain support through Indigenous funding channels, where successful applications result in funding direct to the Indigenous partner organisation.

Third, as technology designers, members of the speech and language technology community may take the lead of scholars in the participatory design community (Winschiers-Theophilus et al., 2010; Simonsen and Robertson, 2012). Technologies already have a reputation of being inappropriate for many Indigenous contexts (Grinevald, 2007, 43), and a new starting point could consider technology as a community practice (Wright and McCarthy, 2004, 108f). Good opportunities for language technologies could be in collaborative contexts, such as the speaker-student-scholar nexus (Bird, 2018, 832ff).

Fourth, we note that language vitality turns on intergenerational transmission. Approaches to sustaining language use need to consider the linguistic pressures on children. One could adopt the mantra:

“it’s the children, stupid” cf. (Wikipedia contributors, 2020). Fortuitously, children will usually be the community segment who are the fastest with technology uptake.

Finally, it may help to accept that the colonial impulse in speech and language technology is not something that can be solved overnight. The challenge is not to “be decolonised” so much as to be participating in a process of decolonisation. This is its own journey of discovery.

5 Conclusion

For a fraction of the world’s languages – perhaps no more than 10% – the dominant ideology is that a language is a communication tool, a public corpus, readily interchangeable with others, raw data for commercial exploitation by algorithms, or hidden behind a “translation layer” (Unbabel, 2019). For the remaining 90%, language tends to be oral, emergent, untranslatable, tightly coupled to a place.

Representatives from the former may approach the latter with a sense of entitlement: to project, to save, to know, to mine. They may be unwilling to hear local aspirations, unable to see how differently language functions in each place. It is simply a given that language loss must be halted, and that technology is up to the task.

Speech and language technologists, myself included, have participated in the “hyperbolic valorization” of endangered languages (Hill, 2002, 120), perpetuating “naive Western assumptions about languages as bounded denotational codes” (Dobrin et al., 2009, 38), leading to “a puristic approach [that] is antithetical to reclamation objectives” (Hermes and Engman, 2017, 71).

The idea of saving languages sits well with our understanding of language as data. It sits well with our belief in technologies as panaceas for social problems (Srinivasan, 2017; Birhane, 2020). It even seems to sit well with our commitment to evaluation. Yet we do not systematically evaluate the claim that speech and language technologies improve the survival prospects for Indigenous or endangered languages. Is this a blind spot?

Indigenous languages are not just acutely under-resourced languages. There is a category distinction, based on profoundly different understandings of the nature and function of language, along with the need to steadily unravel the legacy of colonisation. This calls for a qualitatively different approach, one where we properly theorise the role of speech and language technologies in sustaining language use in a postcolonial world.

* * *

In the Kuwarddewardde “Stone Country” in Northern Australia, I asked my friends Stuart Guymala and Ray Nadjamerrek about their aspirations for *kunwok* “language.” Instead, they spoke about *kunmayali* “knowledge” and *wurdurd* “children.” *Kunmayali ngarribanmarneweykan wurdurd kabarriborlbme munguyihmunguyi* “We’re passing on wisdom for future generations to learn forever and ever.” *Wurdurd ngarribanbukkan kunmayali ba kabarridjordmen wanjh kabarrikukrayekwon* “We’re teaching our children so they will grow up strong.”

Acknowledgements

I am indebted to the Bininj people of Northern Australia for the opportunity to live and work in their community, and particularly to Lois Nadjamerrek, Stuart Guymala, and Dean Yibarbuk for their friendship and guidance over many years. This paper has benefitted from feedback from Mat Bettinson, William Lane, Jennyfer Lawrence Taylor, Brandon Wiltshire, and several anonymous reviewers. This research has been supported by grants from the Australian Research Council entitled *Learning English and Aboriginal Languages for Work*, and the Indigenous Languages and Arts Program entitled *Mobile Software for Oral Language Learning in Arnhem Land*, and covered by a research permit from the Northern Land Council and approvals from the board of Warddeken Land Management and the CDU Human Research Ethics Committee.

References

- Gilles Adda, Martine Adda-Decker, Odette Ambourou, Laurent Besacier, David Blachon, H el ene Bonneau-Maynard, Elodie Gauthier, Pierre Godard, Fatima Hamlaoui, Dmitry Idiatov, Guy-No el Kouarata, Lori Lamel, Emmanuel-Moselly Makasso, Annie Rialland, Sebastian St uker, Mark Van de Velde, Fran ois Yvon, and Sabine Zerbian. 2016. Innovative technologies for under-resourced language documentation: The BULB Project. In *Workshop on Collaboration and Computing for Under Resourced Languages, International Conference on Language Resources and Evaluation*, pages 59–66. ELRA.
- Robert Amery. 2007. Aboriginal language habitat in research and tertiary education. In Gerhard Leitner and Ian Malcolm, editors, *The Habitat of Australia’s Aboriginal Languages: Past, Present and Future*. de Gruyter.
- Antti Arppe, Jordan Lachler, Trond Trosterud, Lene Antonsen, and Sjur Moshagen. 2016. Basic language resource kits for endangered languages: A case study of Plains Cree. *Workshop on Collaboration and Computing for Under-Resourced Languages*, pages 1–8.
- Glenn Auld, Ilana Snyder, and Michael Henderson. 2012. Using mobile phones as placed resources for literacy learning in a remote indigenous community in Australia. *Language and Education*, 26:279–96.
- Marie Ann Battiste and James Youngblood Henderson. 2000. *Protecting Indigenous knowledge and heritage: A global challenge*. Purich.
- Patsy Ngalu Bedford and Siobhan Casson. 2010. Conflicting knowledges: barriers to language continuation in the Kimberley. *Australian Journal of Indigenous Education*, 39:76–86.
- Laurent Besacier, Etienne Barnard, Alexey Karpov, and Tanja Schultz. 2014. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56:85–100.
- Dawn Bessarab and Bridget Ng’andu. 2010. Yarning about yarning as a legitimate method in Indigenous research. *International Journal of Critical Indigenous Studies*, 3:37–50.
- Mat Bettinson and Steven Bird. 2018. Image-Gesture-Voice: a web component for eliciting speech. In *Third Workshop on Collaboration and Computing for Under-Resourced Languages*, pages 1–8.
- Steven Bird and David Chiang. 2012. Machine translation for language preservation. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 125–134.
- Steven Bird and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language*, 79:557–82.
- Steven Bird, David Chiang, Friedel Frowein, Andrea Berez, Mark Eby, Florian Hanke, Ryan Shelby, Ashish Vaswani, and Ada Wan. 2013. The International Workshop on Language Preservation: An experiment in text collection and language technology. *Language Documentation and Conservation*, 7:155–167.
- Steven Bird, Lauren Gawne, Katie Gelbart, and Isaac McAlister. 2014a. Collecting bilingual audio in remote indigenous communities. In *Proceedings of the 25th International Conference on Computational Linguistics*.
- Steven Bird, Florian Hanke, Oliver Adams, and Haejoong Lee. 2014b. Aikuma: A mobile app for collaborative language documentation. In *Proceedings of the Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 1–5. ACL.
- Steven Bird. 2001. Orthography and identity in Cameroon. *Written Language and Literacy*, 4:131–162.
- Steven Bird. 2010. A scalable method for preserving oral literature from small languages. In *Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries*, pages 5–14.
- Steven Bird. 2018. Designing mobile applications for endangered languages. In *Oxford Handbook of Endangered Languages*. Oxford University Press.
- Steven Bird. 2020. Sparse transcription. *Computational Linguistics*, 46(4).
- Abeba Birhane. 2020. Algorithmic colonization of Africa. *SCRIPTed*, 17:389–409.
- Catherine Bow. 2019. Diverse socio-technical aspects of a digital archive of Aboriginal languages. *Archives and Manuscripts*, 47:94–112.
- David Bradley. 2010. Language endangerment and resilience linguistics: Case studies of Gong and Lisu. *Anthropological Linguistics*, 52:123–40.

- Margot Brereton, Paul Roe, Ronald Schroeter, and Anita Hong. 2014. Beyond ethnography: engagement and reciprocity as foundations for design research out here. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1183–86. ACM.
- Mary Bucholtz. 2007. Variability in transcription. *Discourse Studies*, 9:784–808.
- Margaret Carew, Jennifer Green, Inge Kral, Rachel Nordlinger, and Ruth Singer. 2015. Getting in touch: Language and digital inclusion in Australian indigenous communities. *Language Documentation and Conservation*, 9:307–23.
- Jasone Cenoz and Durk Gorter. 2017. Minority languages and sustainable translanguaging: Threat or opportunity? *Journal of Multilingual and Multicultural Development*, 38:901–12.
- Kathy Charmaz. 2014. *Constructing Grounded Theory*. Sage.
- Michael Christie and Helen Verran. 2014. The Touch Pad Body: A generative transcultural digital device interrupting received ideas and practices in Aboriginal health. *Societies*, 4:256–264.
- Michael Christie. 2014. Decolonizing methodology in an Arnhem Land garden. In Beate Neumeier and Kay Schaffer, editors, *Decolonizing the landscape: indigenous cultures in Australia*, pages 57–69. Rodopi.
- Catia Cucchiarini. 1993. *Phonetic transcription: a methodological and empirical study*. Ph.D. thesis, Radboud University.
- Abram De Swaan. 2004. Endangered languages, sociolinguistics, and linguistic sentimentalism. *European Review*, 12:567–80.
- Lise M Dobrin, Peter K Austin, and David Nathan. 2009. Dying to be counted: The commodification of endangered languages in documentary linguistics. *Language Documentation and Description*, 6:37–52.
- Nancy Dorian. 1994. Purism vs. compromise in language revitalization and language revival. *Language in Society*, pages 479–94.
- Nancy Dorian. 2002. Commentary: Broadening the rhetorical and descriptive horizons in endangered-language linguistics. *Journal of Linguistic Anthropology*, 12:134–40.
- Paul Dourish and Scott Mainwaring. 2012. Ubicomp’s colonial impulse. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 133–42.
- Paul Dourish, Christopher Lawrence, Tuck Wah Leong, and Greg Wadley. 2020. On being iterated: The affective demands of design participation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11.
- Ewan Dunbar, Xuan Nga Cao, Juan Benjumea, Julien Karadayi, Mathieu Bernard, Laurent Besacier, Xavier Anguera, and Emmanuel Dupoux. 2017. The Zero Resource Speech Challenge 2017. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*, pages 323–30. IEEE.
- Micha Elsner, Sharon Goldwater, Naomi Feldman, and Frank Wood. 2013. A joint learning model of word segmentation, lexical acquisition, and phonetic variability. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 42–54.
- European Language Resources Association. 2019. LT4All: Language Technologies for All. <https://lt4all.elra.info/en/>, accessed Dec. 2019.
- Joshua A. Fishman. 2001. Why is it so hard to save a threatened language? In Joshua A. Fishman, editor, *Can Threatened Languages be Saved?: Reversing Language Shift, Revisited: a 21st Century Perspective*, pages 1–22. Multilingual Matters.
- Paulo Freire. 1970. *Pedagogy of the Oppressed*. Penguin.
- Donna Gerds. 2010. Beyond expertise: The role of the linguist in language revitalization programs. In Lenore Grenoble and Louanna Furbee, editors, *Language Documentation: Practice and Values*, pages 173–92. John Benjamins.
- Jeff Good. 2018. Reflections on the scope of language documentation. In Bradley McDonnell, Andrea Berez-Kroeker, and Gary Holton, editors, *Reflections on Language Documentation 20 Years after Himmelmann 1998*, number 15 in *Language Documentation and Conservation Special Publication*, pages 13–21. University of Hawai’i Press.

- Jack Goody. 1977. *The Domestication of the Savage Mind*. Cambridge University Press.
- Lenore Grenoble and Lindsay Whaley. 2006. *Saving Languages: An Introduction to Language Revitalization*. Cambridge University Press.
- Colette Grinevald. 2007. Linguistic fieldwork among speakers of endangered languages. *The vanishing languages of the Pacific Rim*, pages 35–76.
- Lev Grossman. 2014. Inside Facebook’s plan to wire the world: Mark Zuckerberg’s crusade to put every single human being online. *Time Magazine*, 184(23).
- Jonathan Grudin. 1994. Computer-supported cooperative work: history and focus. *Computer*, 27:19–26.
- Valérie Guérin. 2008. Writing an endangered language. *Language Documentation and Conservation*, 2:47–67.
- Kenneth Hale, Michael Krauss, Lucille Watahomigie, Akira Yamamoto, and Colette Craig. 1992. Endangered languages. *Language*, 68:1–42.
- Florian Hanke. 2017. *Computer-Supported Cooperative Language Documentation*. Ph.D. thesis, University of Melbourne.
- Mary Hermes and Mel Engman. 2017. Resounding the clarion call: Indigenous language learners and documentation. *Language Documentation and Description*, 14:59–87.
- Mary Hermes and Kendall King. 2013. Ojibwe language revitalization, multimedia technology, and family language learning. *Language Learning & Technology*, 17:125–144.
- Jane Hill. 2002. ”Expert Rhetorics” in advocacy for endangered languages: Who is listening, and what do they hear? *Journal of Linguistic Anthropology*, 12:119–33.
- Nikolaus Himmelmann. 1998. Documentary and descriptive linguistics. *Linguistics*, 36:161–95.
- Nikolaus Himmelmann. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus Himmelmann, and Ulrike Mosel, editors, *Essentials of Language Documentation*, pages 1–30. Mouton de Gruyter.
- Nikolaus Himmelmann. 2018. Meeting the transcription challenge. In *Reflections on Language Documentation 20 Years after Himmelmann 1998*, number 15 in Language Documentation and Conservation Special Publication, pages 33–40. University of Hawai’i Press.
- Leanne Hinton and Kenneth Hale, editors. 2001. *The Green Book of Language Revitalization in Practice*. Academic Press.
- Leanne Hinton. 2001. The Master-Apprentice Language Learning Program. In Leanne Hinton and Kenneth Hale, editors, *The Green Book of Language Revitalization in Practice*, pages 216–226. Academic Press.
- Leanne Hinton. 2015. What counts as a ‘success’ in language revitalization? Seminar, University of British Columbia, <https://ikblc.ubc.ca/leanne-hinton-2/>, Accessed Nov. 2020.
- Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca Grinter. 2010. Postcolonial computing: a lens on design and development. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1311–20.
- Robert Jimerson and Emily Prud’hommeaux. 2018. ASR for documenting acutely under-resourced indigenous languages. In *Proceedings of the 11th Language Resources and Evaluation Conference*, pages 4161–66.
- Mark Johnson and Sharon Goldwater. 2009. Improving nonparameteric Bayesian inference: experiments on unsupervised word segmentation with adaptor grammars. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*, pages 317–25. ACL.
- Anisia Katinskaia, Javad Nouri, and Roman Yangarber. 2017. Revita: a system for language learning and supporting endangered languages. In *Proceedings of the Joint Workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pages 27–35.
- Timothy Kempton and Roger K Moore. 2014. Discovering the phoneme inventory of an unwritten language: A machine-assisted approach. *Speech Communication*, 56:152–66.

- Alexander D King. 2015. Add language documentation to any ethnographic project in six steps. *Anthropology Today*, 31:8–12.
- Sherryl Kleinman and Martha A. Copp. 1993. *Emotions and Fieldwork*. Sage.
- Steven Krauwer. 2003. The Basic Language Resource Kit (BLARK) as the first milestone for the Language Resources Roadmap. *Proceedings of the International Workshop on Speech and Computer*, pages 8–15.
- Don Kulick. 1997. *Language shift and cultural reproduction: Socialization, self and syncretism in a Papua New Guinean village*. Cambridge University Press.
- Clare Land. 2015. *Decolonizing Solidarity: Dilemmas and Directions for Supporters of Indigenous Struggles*. Zed Books.
- Paul Lewis and Gary Simons. 2010. Assessing endangerment: Expanding Fishman’s GIDS. *Revue Roumaine de Linguistique*, 55:103–120.
- Paul Lewis and Gary Simons. 2016. *Sustaining Language Use: Perspectives on Community-Based Language Development*. SIL International.
- Anant Maheshwari, Léo Bouscarrat, and Paul Cook. 2018. Towards language technology for Mi’kmaq. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pages 4139–43. ELRA.
- Doug Marmion, Kazuko Obata, and Jakelin Troy. 2014. *Community, Identity, Wellbeing: The Report of the Second National Indigenous Languages Survey*. Australian Institute of Aboriginal and Torres Strait Islander Studies, Canberra.
- Lutz Marten and Malin Petzell. 2016. Linguistic variation and the dynamics of language documentation: Editing in ‘pure’ Kagulu. In Mandana Seyfeddinipur, editor, *African Language Documentation: New Data, Methods and Approaches*, number 10 in Language Documentation and Conservation Special Publication, pages 105–129. University of Hawai’i Press.
- Francesca Merlan. 2009. Indigeneity: Global and local. *Current Anthropology*, 50:303–33.
- Alexis Michaud, Oliver Adams, Trevor Cohn, Graham Neubig, and Séverine Guillaume. 2018. Integrating automatic transcription into the language documentation workflow: experiments with Na data and the Persephone Toolkit. *Language Documentation and Conservation*, 12:481–513.
- Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy and Technology*, pages 1–26.
- Robert Moore. 2006. Disappearing, Inc.: Glimpsing the sublime in the politics of access to endangered languages. *Language and Communication*, 26:296–315.
- Mark Moran. 2016. *Serious Whitefella Stuff: When solutions became the problem in Indigenous affairs*. Melbourne University Press.
- Ulrike Mosel. 2006. Fieldwork and community language work. In Jost Gippert, Nikolaus Himmelmann, and Ulrike Mosel, editors, *Essentials of Language Documentation*, pages 67–85. Mouton de Gruyter.
- Markus Müller, Jörg Franke, Alex Waibel, and Sebastian Stüker. 2017. Towards phoneme inventory discovery for documentation of unwritten languages. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 5200–04. IEEE.
- David Nathan and Meili Fang. 2014. Re-imagining documentary linguistics as a revitalization-driven practice. In Mari Jones and Sarah Ogilvie, editors, *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*, pages 42–55. Cambridge University Press.
- M. Eleanor Nevins. 2013. *Lessons from Fort Apache: Beyond language endangerment and maintenance*. Wiley.
- NHMRC. 2018. *Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders*. National Health and Medical Research Council.
- Robyn Ober. 2017. Kapati time: Storytelling as a data collection method in Indigenous research. *Learning Communities: International Journal of Learning in Social Contexts*, 22:8–15.
- Elinor Ochs. 1979. Transcription as theory. *Developmental Pragmatics*, 10:43–72.

- Ray Pawson. 2006. *Evidence-Based Policy: A Realist Perspective*. Sage.
- Ray Pawson. 2013. *The Science of Evaluation: a Realist Manifesto*. Sage Publications.
- Bernard Perley. 2012. Zombie linguistics: Experts, endangered languages and the curse of undead voices. *Anthropological Forum*, 22:133–49.
- Aidan Pine and Mark Turin. 2017. Language revitalization. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.
- Maïa Ponsonnet. 2019. *Difference and Repetition in Language Shift to a Creole: The Expression of Emotions*. Routledge.
- Keren Rice. 2009. Must there be two solitudes? Language activists and linguists working together. In Jon Reyhner and Louise Lockhard, editors, *Indigenous language revitalization: Encouragement, guidance, and lessons learned*, pages 37–59. Northern Arizona University.
- William Samarin. 1984. The linguistic world of field colonialism. *Language in Society*, 13:435–53.
- Racquel-María Sapién. 2018. Design and implementation of collaborative language documentation projects. In *Oxford Handbook of Endangered Languages*, pages 203–24. Oxford University Press.
- Corinne Seals and Vincent Olsen-Reeder. 2020. Translanguaging in conjunction with language revitalization. *System*, page 102277.
- Amartya Sen. 1999. *Development as Freedom*. Oxford University Press.
- Jesper Simonsen and Toni Robertson, editors. 2012. *Routledge International Handbook of Participatory Design*. Routledge.
- Linda Tuhiwai Smith. 1999. *Decolonizing Methodologies*. Zed Books, 2nd edition.
- Margaret Somerville and Tony Perkins. 2003. Border work in the contact zone: Thinking indigenous/non-indigenous collaboration spatially. *Journal of Intercultural Studies*, 24:253–66.
- Alessandro Soro, Margot Brereton, Jennyfer Lawrence Taylor, Anita Lee Hong, and Paul Roe. 2016. Cross-cultural dialogical probes. In *Proceedings of the First African Conference on Human Computer Interaction*, pages 114–25. ACM.
- Ramesh Srinivasan. 2017. *Whose Global Village?: Rethinking how Technology Shapes Our World*. NYU Press.
- Susan Leigh Star and James R. Griesemer. 1989. Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19:387–420.
- Susan Leigh Star. 2007. Living grounded theory: Cognitive and emotional forms of pragmatism. *The Sage Handbook of Grounded Theory*, pages 75–94.
- Susan Leigh Star. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science, Technology and Human Values*, 35:601–617.
- Tonya Stebbins, Kris Eira, and Vicki Couzens. 2017. *Living Languages and New Approaches to Language Revitalisation Research*. Routledge.
- Katy Steinmetz. 2018. One of the world’s 7,000 languages dies every three months: Can apps help save them? *Time Magazine*. <http://time.com/5417035/technology-endangered-languages/>, Accessed Nov. 2020.
- Jennyfer Lawrence Taylor, Alessandro Soro, Margot Brereton, Anita Lee Hong, and Paul Roe. 2016. Designing evaluation beyond evaluating design: measuring success in cross-cultural projects. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pages 472–77.
- Jennyfer Lawrence Taylor, Wujal Wujal Aboriginal Shire Council, Alessandro Soro, Paul Roe, and Margot Brereton. 2019. A relational approach to designing social technologies that foster use of the Kuku Yalanji language. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 161–72.
- John Taylor. 2008. Indigenous peoples and indicators of well-being: Australian perspectives on United Nations global frameworks. *Social Indicators Research*, 87:111–26.

- Dennis Tedlock. 1983. *The Spoken Word and the Work of Interpretation*. University of Pennsylvania Press.
- Unbabel. 2019. Unbabel raises \$60m Series C to become the world’s translation layer. <https://unbabel.com/news/unbabel-raises-60m-series-c-to-become-the-translation-layer-of-the-internet>, Accessed Nov. 2020.
- UNESCO Ad Hoc Expert Group on Endangered Languages. 2003. *A Methodology for Assessing Language Vitality and Endangerment*. UNESCO. <http://www.unesco.org/new/en/culture/themes/endangered-languages/language-vitality/>, Accessed Oct. 2020.
- United Nations. 2010. *State of the World’s Indigenous Peoples*. United Nations Department of Public Information.
- Helen Verran and Michael Christie. 2007. Using/designing digital technologies of representation in Aboriginal Australian knowledge practices. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, 3:214–27.
- Helen Verran, Michael Christie, Bryce Anbins-King, Trevor Van Weeren, and Wulumdhuna Yunupingu. 2007. Designing digital knowledge management tools with Aboriginal Australians. *Digital Creativity*, 18:129–42.
- Marco Vetter, Markus Müller, Fatima Hamlaoui, Graham Neubig, Satoshi Nakamura, Sebastian Stüker, and Alex Waibel. 2016. Unsupervised phoneme segmentation of previously unseen languages. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, pages 3544–48.
- Michael Walsh. 2005. Will indigenous languages survive? *Annual Review of Anthropology*, 34:293–315.
- Maggie Walter and Michele Suina. 2019. Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22:233–43.
- Monica Ward and Josef Genabith. 2003. CALL for endangered languages: Challenges and rewards. *Computer Assisted Language Learning*, 16:233–58.
- Monica Ward. 2018. Qualitative research in less commonly taught and endangered language CALL. *Language Learning and Technology*, 22:116–32.
- Douglas Whalen, Margaret Moss, and Daryl Baldwin. 2016. Healing through language: Positive physical health effects of indigenous language use. *F1000Research*, 5.
- Wikipedia contributors. 2020. It’s the economy, stupid — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/It%27s_the_economy,_stupid, Accessed Nov 2020.
- David P Wilkins. 2000. Even with the best of intentions...: Some pitfalls in the fight for linguistic and cultural survival (one view of the Australian experience). *As linguas amazonicas hoje: the Amazonian languages today*, pages 61–83.
- Lorna Williams. 2019. Wa7 szum’in’stum’ ti nqwelutenlhkalkha: Technology and Indigenous language revitalization, recovery and normalization. Keynote address at International Conference Language Technologies for All (LT4All): Enabling Linguistic Diversity and Multilingualism Worldwide. <https://en.unesco.org/LT4All>, Accessed Jan. 2020.
- Heike Winschiers-Theophilus, Shilumbe Chivuno-Kuria, Gereon Koch Kapuire, Nicola Bidwell, and Edwin Blake. 2010. Being participated: a community approach. In *Proceedings of the 11th Biennial Participatory Design Conference*, pages 1–10. ACM.
- Peter Wright and John McCarthy. 2004. *Technology as Experience*. MIT Press.
- Ghil’ad Zuckermann. 2020. *Revivalistics: From the Genesis of Israeli to Language Reclamation in Australia and Beyond*. Oxford University Press.