

Dylan Scott Low

National Taiwan Normal University, Taiwan

Isaac Mcneill

Mahidol University, Thailand

Michael James Day

Xi'an Jiaotong-Liverpool University, China

ENDANGERED LANGUAGES: A SOCIOCOGNITIVE APPROACH TO LANGUAGE DEATH, IDENTITY LOSS, AND PRESERVATION IN THE AGE OF ARTIFICIAL INTELLIGENCE

Summary. Sociolinguists suggest language death entails significant cultural, personal, and ecological loss. Socio-cultural and socio-political factors exacerbate language erosion and encourage supplantation by another more dominant language. Hence, we ask: what are the sociocognitive principles which make language death hurtful and symbolic? Within this article, we attempt to outline a sociocognitive account of language death, situating the Hallidayan perspective of language as a “social-semiotic” system alongside a Cognitive Linguistic approach. We further contextualise language as inseparable from culture, drawing insight from the sociological thought of Bourdieu. We contend that language death entails psychological trauma, representing the destruction of cultural genealogy and the loss of knowledge intrinsic to personal self-imagery and identity. To this end, we present a case study of the Māori languaculture in Aotearoa (New Zealand), tracing the impact of colonialism and marginalisation to current efforts and ambitions to ensure the languacultural survival of Māori and reclaim space in Aotearoa as a respected knowledge system and means of expression, particularly in the socio-technical age of artificial intelligence (AI) and the Web. We argue that our analysis bodes practical implications for language maintenance and revitalisation, concluding that sociolinguistic practitioners should consider a socio-cognitivist as well as socio-technical paradigm for language intervention. In closing, we discuss leveraging AI technologies towards language heritage, archival, and preservation to limit the destructive impact of the death of a language.

Keywords: language death; culture; identity; cognition; sociolinguistics; artificial intelligence.

Introduction

Languages can die. Language death, which sociolinguists regard as when a language loses its native speakers and ceases to be spoken, is a traumatic experience for its cultural community, usually ethno-cultural minorities, with significant social and cultural-ecological implications (Crystal, 2002). It is estimated that 50–90% of endangered languages will disappear in the next

100 years, taking history, communal connectivity, and heritage along with them (Nuwer, 2014). Yet, there are theoretical implications as well, which extend across an interdisciplinary mindscape. Adopting the functionalist idea of language as a “social-semiotic” system (Halliday & Hasan, 1985), language death entails two significant losses: the loss of cultural heritage and a unique system of thought expression. These tend to be taken separately, but they are as intertwined in language death as they are in life; culture is the accrued capital that fuels our symbolic communication (Bourdieu, 1972–1977). Hence, contemporary threats to the heritage of cultural bodies, accompanied by the impact of social transformation as a result of globalisation, are reshaping the verbal expression of culture.

From a cognitive perspective, linguistic speech acts involve the mapping of meaning to symbols, a semiotic process (Langacker, 2008; Lamb, 1999). The cognitive processes underlying this cannot be decoupled from experiences that give rise to our meanings (Lakoff & Johnson, 1980). To this end, Cognitive Linguistics is an interdisciplinary endeavour where language arises from general cognitive competencies. Yet, its parent discipline, Cognitive Science, has been undergoing a paradigm shift, from a classical computationalist approach envisioning the mind as a modular algorithmic machine, to a more dynamic approach which situates the mind, broadly, as shaped by context (Varela et al., 1991). So, there is now greater recognition that the mind, and thus, language, operates in the context of culture and environment, a socio-technical perspective (Low et al., 2020). Indeed, whilst Cognitive Linguistics has been traditionally disconnected from Sociolinguistics, work uniting the two has begun to emerge as “Cognitive Sociolinguistics” (cf. Geeraerts et al., 2010). Yet, language death and erosion issues have remained understudied from the cognitive approach.

Hence, we attempt to outline a cognitive approach to language death. We contend that language death, as an event, is symbolic. Furthermore, we highlight the importance of emergent technologies such as artificial intelligence in preventing language death. Consequently, we lay the foundation for a framework where a language death event itself has meaning, or *cultural capital*, to borrow from Bourdieu (1972/1977; 1979/1984). From the perspective of the community whose heritage language has died, this

cultural capital possesses a negative value, potentially traumatic and painful. Consequently, when a knowledge system is faced with extinction, we contend that effective use of technologies may form a solution to preserve it, as well as its wider culture. Hence, the question we ask is: What are the cognitive principles that make language death traumatic, and how can such destruction be mitigated?

Sociological Considerations: What is Language?

Language serves major social functions. Indeed, society can hardly function without language. There is a consensus amongst linguists that language is unique and exclusive to human beings, separating us intellectually from animal species. Where the consensus ruptures, however, is on the issue of what exactly language is. The Chomskyan tradition contends that language is a system of formal rules innate to our biology, and that the goal of linguists should be to uncover these universal principles, which constitute a mental faculty autonomous from other cognitive, and social, functions. Alternatively, the functionalist paradigm understands language as a tool for social, communicative, and even survival purposes. In this view, language is very much subject to social pressures and developed out of these pressures through our species' evolution (Joseph, 2018). Additionally, this entails the implication that language cannot be autonomous from other realities of human life, such as culture, wider cognitive systems, wider social systems, and brain structure (Agar, 2019; Halliday & Hasan, 1985; Lamb, 1999).

Joseph (2018, p. 4) levies significant criticism against the Chomskyan conception of language, criticising Chomsky for attempting to perpetuate a "...façade of scientific certainty" about Linguistics by manufacturing the concepts of *I-language*, or "internalised language", and *E-language*, or "externalised language", to disconnect the discipline from social phenomena. Borrowing from Latour (1993/1991), Joseph identifies language as a hybrid concept between "nature" and "society" (2018, p. 10). To use Latourian parlance, the denial of the social dimensions of Linguistics is a *work of purification*, which we may juxtapose against a *work of translation*, meaning the conceptualisation of "hybrids of nature and culture" within the scientific

theory (Latour, 1993–1991, p. 10). It is argued that works of translation are more scientifically productive (Joseph, 2018).

Unfortunately, language death seems to have received a different work of purification, one that has undercut its connection to the mind and body. Crystal (2002) provides five reasons why language death is something to be cared about: because we need diversity, because languages express identity, because languages are repositories of history, because languages contribute to the sum of human knowledge, and because languages are interesting in themselves. Yet, Crystal's reasons would hardly be convincing to the Chomskyan linguist; for Chomsky, these all fall in the "artificial" and "not very interesting" realm of E-language (Chomsky, 1986, p. 26). Crystal's reasons are thus susceptible to Chomsky's critique of E-language. These phenomena, such as a language's suppression, disuse, and death, are suggested by Chomsky to be arbitrary constructs built on the whims of socio-political factors external to Linguistics and "understood independently of the mind/brain" (Chomsky, 1986, p. 20).

Whilst a well-known political activist, Chomsky (1988, p. 394) maintains that "Surely one cannot simply deduce social and political consequences from any insights into language". Yet, to purify the language of either social or bio-cognitive entailments is, in our view, counter-productive in defining language itself. Consequently, here we present our challenge; we reject the idea that "E-language", if such a label is to be used, is scientifically uninteresting. Whilst we are more sympathetic to the sociolinguistic position, we contend we must also call for more consideration of cognitive operations in (socio-)linguistic practice. To this end, we suggest that to better understand the cognitive principles behind language, and language death, we must first situate language in context. Hence, a work of translation is necessary, and we engage this by outlining a sociological approach to language, starting with a social-semiotic conception.

Work of Translation: Language as a Social-semiotic System

Language and culture are inseparable. Within Systemic Functional Linguistics (SFL), Halliday argues for language as a "social-semiotic" system with which

we use to construe our experience and enact social relationships, alongside other systems such as art, ceremony, and even cuisine (Halliday & Matthiessen, 2004). One may identify a Saussurean aspect here; for de Saussure, Linguistics formed a sub-branch of a more general study of “the life of signs within society” (de Saussure, 1916–1959, p. 16). Other thinkers further emphasise the unity between linguistic and cultural systems, such as Agar (2019; 1993), who champions the term *languaculture*. Similarly, Halliday (2016, p. 115) contends that, as an example, the English language, through its colonial past and present international status, has become adapted to the “meaning styles of its new speakers”, so culturally contingent contexts of thought and expression. Such cultural “meaning styles” can be interpreted as an aspect of Bourdieu’s concept of *habitus*.

Halliday’s perspective on language, then, gives primacy to semantic structures. Yet, Halliday also links semantics to culture, summarising language as “...one among a number of systems of meaning that, taken all together, constitute human culture” (Halliday & Hasan, 1985, p. 2). Chomsky draws a distinction between linguistic *competence*, which refers to a person’s ability to generate linguistic material, and linguistic *performance*, which corresponds to Halliday’s notion of *instance*, the production of text and meaningful cultural products; for example, speech, books, social media posts, or instant messages (Chomsky, 1986; Halliday & Hasan, 1985). Distinct from Chomsky’s understanding, Halliday’s concept is tied to social interaction and enactment; Halliday and Matthiessen (2004) qualify “text” broadly as an instance of language in some medium that is comprehensible to someone who understands the language. Derrida (1972–1982) further qualifies text as necessarily *iterable*, meaning it is repeatably reinterpretable by some interlocutor; when a text is no longer acknowledged as a meaningful unit, it loses its iterability and ceases to be text.

An iterable text is also necessarily subjected to reinterpretation each time it is processed by an interlocutor; our understanding of a text is always construed in terms of our own conceptual resources (Derrida, 1972–1982; Lakoff & Johnson, 1999). Within Cognitive Linguistics (CL), language use is embedded in the mind’s wider cognitive competencies. Language expresses thought, organised as meaning, and to an extent, reveals certain common

patterns by which we think (Lakoff & Johnson, 1999). The compatibility between SFL and CL is emphasised by Butler (2013), who highlights Halliday's notion of language as a *meaning-making resource*. The "systemic" part of Halliday's SFL thus refers to the conception of language as consisting of systems of potential meaning, whilst both SFL and CL agree that language construes human experience (Halliday & Matthiessen, 2004; Langacker, 2008; Lakoff & Johnson, 1999). Yet, SFL also emphasises that such construal of experience, via language, always enacts some social action.

This refers to what is termed the *interpersonal metafunction* of language, and grounds the "social-semiotic" view (Halliday & Matthiessen, 2004). The interpersonal metafunction works alongside the *ideational* and *textual* metafunctions to create coherent, meaningful, and contextually relevant language. All instances of language simultaneously perform these triadic metafunctions (Halliday & Matthiessen, 2004). In brief, the ideational metafunction is responsible for the construal of experience and logic into lexicogrammatical structure, and the textual metafunction is responsible for organising linguistic units into coherent and flowing orders of discourse (Halliday & Matthiessen, 2004; Foucault, 1966–1970). The interpersonal metafunction denotes those grammatical systems which help the speaker establish their attitudes, position, and relation to other actors; consequently, it also acts as a contextual constraint on what can be said by delineating the appropriateness of a speaker's possible utterances.

Such relationships entail an aspect of power, as well as economic exchanges of information (Foucault, 1966/1970; Bourdieu, 1982–1991). The "social-semiotic" conception of language, as developed within the SFL paradigm, thus offers a socially-grounded perspective for language issues. Despite Halliday's commitment to Linguistics as "a branch of sociology" and the belief that "...there is no need to interpose a psychological level of interpretation" to language (Halliday, 1978, pp. 38–39), such a "psychological" extension of SFL is precipitated by SFL's compatibility with CL theories of experiential construal. What SFL adds of value to cognitivist approaches to language is its "social-accountability" (Matthiessen, 2012). This was a deficiency of Chomsky's linguistic cognitivism, though it later inspired Bourdieu's response in his theory of linguistic practice, emphasising language

in social context (Salö, 2019). Bourdieu offers a perspective to unite the “social-semiotic” conception with Cognitive Sociology, and adds value to Linguistics by outlining a cogent framework for situating language in society.

Bourdieu’s Theory of Linguistic Practice

Words have power. Language is thus a vehicle to enact complex interpersonal relationships and construct social realities (Halliday & Matthiessen, 2004). Bourdieu (1979–1984, p. 101) condenses his sociological theory into a simple formula: [(habitus)(capital)] + field = practice. Bourdieu also sought to fuse the study of language with the study of society. According to Bourdieu, *habitus* refers to “...a subjective but not individual system of internalised structures, schemes of perception, conception, and action common to all members of the same group or class” (1972–1977, p. 86). Thus, *habitus* refers to the collection of dispositions that guide the actions and reactions of an individual in general correlation with their social group. For example, picture a Western tourist in Thailand during *Songkran*, the Thai New Year. Holding a super-soaker and a beer, the tourist had come to the country for the “world’s biggest water fight”, as media often markets Songkran. Consider then, during the same festival, an 80-year-old Thai woman visiting her local Buddhist temple to offer food to monks and give blessings in a traditional ritual.

These two individuals, situated in the same geographical setting, have different dispositions, formulated from divergent experiences owing to their different histories. In discussing language, Bourdieu proposes a certain *linguistic habitus*. *Habitus* are formed through social inculcation, and thus serve as a mechanism for the reproduction of culture; a person’s *habitus* is built from their early childhood interactions with their parents, and later through their education and other social activities (Bourdieu, 1977–1972). Similarly, the linguistic *habitus* is developed through early childhood acquisition of one’s native language, dialect, and meaning style (Bourdieu, 1982–1991). This account of acquisition is similar to Lakoff and Johnson’s (1980; 1999) arguments for embodied cognition, which suggested that pre-linguistic thought was structured by experience with the external world.

Bourdieu also recognised the embodiment of language, principally

through its link to the human vocal tract and mouth, which enable speech (Bourdieu, 1982–1991). Scovel (1988, p. 14) independently extended this idea, claiming that the emergence of language in human evolution has to be credited to the lateral development of our unique vocal tract system, as the “organ of speech”, alongside the brain, as the “organ of language”. However, whereas Scovel’s account seems to neglect non-speech-based linguistic systems, such as signed languages amongst the deaf, Bourdieu would likely have noted that spoken languages and signed languages are distinguished by different modes of embodiment (one uses oral-respiratory apparatus while the other relies mainly on hand-based expression), and that this entails different access to symbolic capital. Such modes of embodiment are themselves symbolic; despite the fact that signed languages do not lack any cognitive functionality possessed by spoken languages, many cultures associate signed languages with disability.

Symbolic capital refers to elements that, like currency, can be exchanged for benefits (Bourdieu, 1972–1977; 1979–1984). The lectureship title *Ajarn* in Thailand is an example of culture-specific symbolic capital (Day et al, 2021). For a more general example, a degree from a brand name school possesses more symbolic capital than one from a less famous school. On the other hand, *cultural capital* refers to the collection of one’s cultural knowledge which informs their habitus, consequently guiding their successful navigation of cultural context. For example, to fully appreciate the poem *Beowulf*, one would need to possess certain forms of cultural capital, such as sufficient knowledge of Old English and historical knowledge. Knowledge of Old English, then, would constitute a subtype of cultural capital called *linguistic capital* (Bourdieu, 1982–1991).

This linguistic capital refers to the linguistic resources we access to function in certain linguistic fields, which are the contexts in which certain configurations of “meaning styles”, to borrow from Halliday, are dominant. To illustrate, a person who speaks Singapore English (SE) may not appreciate much an American English (AmE) pun where the word “duty” [ˈduːri] is humorous because the word sounds similar to “doody” [ˈduːri] in AmE. Yet, in SE, “duty” [ˈdjuːti] and “doody” [ˈduːdi] are not homophonic, and “doody” is a rare word. This situation depicts an intercultural conflict, what Agar (1993)

calls a *rich point*, a divergence between different *languacultures*, which are Bourdieusian fields. Here, the meaning of the pun is lost because of differences in the linguistic capital of SE and AmE speakers.

Constructing Language Death: Neurophenomenology and A Case from Te Reo Māori

Within Sociolinguistics, the more common term for what we have been calling “field” is “ecology” (cf. Crystal, 2002). This choice of terminology is significant because it is an employment of metaphor; in particular, that *language is an organism*. The term “language death”, then, is congruent to this metaphor and reinforces it. Hence, “language death”, as well as the synonymous phrase “language extinction”, invoke the above-mentioned *language is an organism* metaphor, whilst the related term “language erosion” instantiates the metaphor *language is a physical object* (Lakoff & Johnson, 1980). The choice of these specific metaphorical construals amongst the academic linguistic community is arguably underpinned by a discursive decision to attach to language, as an entity, the connotations that, like living organisms, language is real and has ecological value worth protecting.

This idea that languages constitute natural real-world entities, according to Lamb (1999), is fallacious. This is because each person possesses a unique configuration of their linguistic habitus, a subset of their complete mental model of perceived reality (Lamb, 2020). Thus, what is typically called “English” is but a set of abstract generalisations across heterogeneous unique and personal mental models of “English” by speakers around the world. “English”, then, does not exist as a natural entity independent of its speakers. Rather, it is embedded in the linguistic habitus of its speakers; a monolingual English speaker operates according to their personal understanding of the norms and conventions of “English”. Each English speaker may be described as having their own ways of linguistically expressing themselves, with “English” merely being a label for a group of similar patterns of expression. What sociolinguists call “language death”, then, must be construed not as the loss of discrete natural entities but as the loss of the knowledge of *ways of expression*.

Yet, just because languages are not living organisms does not make their losses any less traumatic. Like culture, natural languages owe their continuous existence to human behaviours (Low et al., 2020). Parents raise their children in some language A, and the children grow up to speak language A. The repetition of this process across generations allows language A, as a languacultural knowledge system, or habitus, to have a continuous genealogy. For Lamb (1999) and Feldman (2006), we can model linguistic knowledge as patterns of neural network circuitry, susceptible to biological rewiring, or *neuroplasticity*. Malabou (2012) characterises neuropathological diseases such as Alzheimer's as "explosive" trauma that erases selfhood; as the neural connections which make up the self deteriorate, the original self ceases to exist whilst a new, less recognisable self emerges. Language death, then, involves the retirement of patterns associated with a language from the cultural bio-genealogy of a population.

To the outside observer, then, language A may simply be no longer spoken by its people, perhaps replaced by language B. Especially problematic, is when this happens as supplantation by a politically powerful external culture, as in colonial or neo-colonially driven language death. Phenomenologically, the community associated with language A experiences a trauma akin to Malabou's description: the self which language A was a part of has been erased, replaced by a new unfamiliar self speaking some imposed language B. Language A's genealogy is ended, barring unreachable traces of memory, as with neurodegeneration. As Erikson's (1968) psychoanalysis of self-identity expounds, people coordinate multiple senses of self, which form the individual's *composite Self*. A lost language represents a constituent of the composite Self, then; language death proves psychologically harmful because it damages the ability of the individual to access their sense of self associated with that language. Such damage may consequently lead to *identity crisis*, in which the individual struggles with "...not just a matter of contradictory self-images... but a central disturbance dangerous for the whole ecological interaction of a mind organism [the individual] with its 'environment'" (Erikson, 1970, p. 749). Thus, language death, especially by marginalisation, creates a pathway to social dysfunction.

Viewing language death as psychological trauma, furthering

dysfunction, which may lead to crisis, echoes Du Bois' (1899) observations that marginalised groups, such as African Americans in late-nineteenth century Philadelphia, tend to experience little socioeconomic success not because of inherent deficiencies but rather psychological harm inflicted by more dominant groups. However, marginalisation is complex, and language death does not always entail dysfunction; the Baba Malay-speaking Peranakan Chinese of Singapore, for example, experienced a trend of socioeconomic advancement in the early twentieth-century British colonial era owing to their early adoption of English-medium education, rather than Chinese or Malay, for their children (Chew, 2013). Rather, the abandonment of Baba Malay, now a moribund language, might be said to represent an unconscious effort to preserve social function, at the expense of cultural damage. The complexity of marginalisation, languacultural loss, and identity entails that such issues take vastly different forms and meanings in different contexts. To illustrate, we explore a case study from a specific languaculture experiencing loss, namely the Māori people of Aotearoa (New Zealand).

Languacultural Loss, Colonialism, Assimilation, And the Revitalised Recovery of Te Reo Māori

Te reo Māori (te reo) is the Indigenous language of Aotearoa, or New Zealand (NZ). In the 1800s, te reo was widely spoken. During this period, European settlers used te reo in social networks that facilitated trade within NZ's majority Māori population. The use of te reo diminished after *Pākehā* (NZ Europeans) became the majority in the 1860s and English took over as the primary language for commerce, education, and communication. The land wars of the late 1800s saw massive confiscations of Māori land, when British soldiers and *iwi* (tribes) allies used a scorched earth policy and the government eventually took control of NZ by force. By 1900 it was estimated that Māori, who had numbered in the hundreds of thousands prior to settlers arriving, had a population of only 42,000 (Reedy, 2000, p. 157). It was predicted that Māori would become extinct as a people. The near-death of Māori as a people was symbolic of the subsequent near-death of their languaculture. Indeed, the colonial suppression of te reo is well documented by scholars (Reedy, 2000;

Tau, 2001; O'Toole, 2020).

Two significant factors that caused the decline in the use of te reo were the government assimilation policies which prioritised English and penalised the use of Māori in education institutions, and the urbanisation of Māori in the mid-twentieth century. O'Toole (2020) notes that a 1930 survey of children attending Native schools showed that approximately 96.6% spoke te reo at home. O'Toole also observed that intergenerational language transmission declined through their urban disconnection to the *rohe* (tribal lands). Indeed, just thirty years later in 1960, only 26% used te reo exclusively at home, "...and by 1979, te reo was considered moribund" (O'Toole, 2020, p. 199). However, in the 1970s, Māori began to restore their identity. The Māori Language Act was passed in 1987, making te reo an official language of NZ. This legislation was replaced by the Māori Language Act 2016 and the formation of Te Mātāwai in 2018, which created a responsibility structure for protection of te reo between the NZ government and Māori; progress has been made by Kohanga Reo, Kura Kaupapa, Wānanga, and other educational institutions to revitalise te reo, including its *iwi* (tribal) dialects.

Revival of the te reo is symbolic of Māori resilience, but the trauma of colonial domination has devastating psychological consequences. Radio New Zealand (2019) reported the "suicide rate among Māori men rose to almost 32 per 100,000 in 2016 – more than double the non-Māori male rate." The revitalisation of te reo through government policy, educational institutions, media and at a broader social level involving both Māori and non-Māori is now a priority. This reaffirms the language, granting more power than seen previously. There are significant signs of increasing use of te reo in the home, in education, in government and in businesses; but the long term use of te reo within a sustainable population remains doubtful.

The political discourse of ideas, knowledge, and language are essential indicators of dominance by some groups over others (Bourdieu, 1982–1991; Foucault, 1980). The colonial imposition of British cultural hegemony saw the NZ government enforce English as the dominant language whereas te reo and Māori culture were deemed inferior to that of Pākehā. The "superiority" of Western knowledge was applied by the NZ government as a method to accumulate power and dominance over Māori. Te reo has always been central

to Māori political expression and identity as *tangata whenua* (people of the land). For Māori, te reo is crucial to the socio-political struggle for *mana motuhake* (self-determination) and social justice; language survival is a form of cognitive justice. Cognitive justice seeks to address the dominance of Eurocentric knowledge and asserts that the plurality of non-Western knowledge has a right to coexist with Western knowledge (Leibowitz, 2017). What it seeks is a space for dialogue between knowledge systems in order to achieve social justice. However, in some institutions there is resistance to more culturally inclusive curricula due to the wrong belief that Western knowledge is superior. In 2021, *mātauranga Māori* (Māori knowledge) was dismissed as “not science” in a letter by some top academics at the University of Auckland (Dunlop, 2021).

Corporate use of Māori languaculture such as the *haka*, which has been commercialised by the All Blacks rugby team and pop artist Lorde’s 2021 release of an EP in te reo, bring up problematic questions as to whose responsibility it is to guide the revitalisation of te reo. Smith (2012, p. 4) makes a compelling case for decolonising and indigenising approaches within academic institutions, to counter the destructive legacy of imperialism, remarking that “...increasing numbers of indigenous academics and researchers have begun to address social issues within the wider framework of self-determination, decolonization and social justice.” In their discussion of te reo and NZ national identity, O’Toole (2020, p. 209) states that, “As more non-Māori engage with te reo in ways meaningful to them, the value and place of Māori culture in New Zealand society shifts.” Regardless of its revitalisation, te reo is still in danger of language death. It is listed as “vulnerable” by UNESCO. The Ministry for Ethnic Communities (2021) lists just 3% of New Zealand’s population as te reo speakers, according to the 2013 NZ Census. Barrett-Walker et al. (2020) concluded that current learning rates of te reo show that it is on a path towards extinction and that proficient speakers must focus on teaching young Māori as a priority before non-Māori. However, whilst significant progress has been made, much more needs to be done to protect, promote, and preserve te reo if it is to continue to be passed down as a knowledge-system and means of expression. Consequently, the Ministry of Māori Development, Te Puni Kōkiri (2019, pp. 11–13), has put forth “audacious goals” which state that:

“By 2040, 85 per cent (or more) of New Zealanders will value te reo Māori as a key part of national identity. By 2040, one million (or more) New Zealanders will have the ability and confidence to talk about at least basic things in te reo Māori. By 2040, 150,000 Māori aged 15 and over will use te reo Māori at least as much as English.”

Whilst “audacious”, such goals underscore a desire for a bright future for te reo. This is especially true in a digital era, when language is changing rapidly, with new discourse unfolding and evolving constantly. Yet, technology, especially the Web, needs to be utilised more to preserve te reo and its dialects. It is crucial for archiving *taonga* (cultural treasures) and *tikanga Māori* (customary practises) that may be lost. Furthermore, socio-cultural factors such as anti-te reo sentiment in the Pākehā population can proliferate both online and offline, and revitalisation efforts must tackle this. However, also significant, Māori need to overcome the language trauma of te reo’s near-death and its psychological barriers. For example, Morrison (2019, p. 9) suggests that learning te reo should include fun and *whanaungatanga* (family) based activities to address “...common barriers such as whakamā, or embarrassment and language trauma from past experiences...”. In other words, revitalisation must encourage resistance to the colonially-shaped languacultural field which devalues te reo. With social media technologies playing a larger role in the evolution of culture, younger generations of Māori fluent in digital technology may be able to enact a cultural shift by using te reo online.

In discussing the need for the Ngāi Tahu iwi to adapt, Tau (2001, p. 150) states that “As devastating as it was, the arrival of the Pākehā did one thing; it showed our ancestors that there was a world beyond these shores and that if we do not actively engage with the rest of the world Ngāi Tahu will become an artefact that exists by itself in the ‘intangible void’.” Undoubtedly, Māori must continue to adapt and look ahead to the future with technology as a tool for the emancipation of te reo from language death. There are many important social, cultural, and technical traditions embedded within many cultures that could just have easily disappeared. Yet, part of the problem is that agendas towards preservation usually arise from the efforts of those communities whose survival is under threat. This comes at the cost of diminishing returns due to their limited resources, adding a form of cultural

shock driven by the complexity of marginalisation, languacultural loss, and political diminishment, especially in authoritarian settings (Mcneill, 2021).

Interdisciplinary Implications: Language AI Preservation and Revitalisation

The question, then, is of external investment, given cultural assimilation and colonisation originating from Western settings, which emerge dominant. Our implication is that technology, or socio-technicality, can be used more effectively to preserve and revitalise languacultures, such as the Māori traditions described above. Artificial intelligence (AI), namely the incorporation of large data sets and analytical instruments that enable machines to undertake complex modelling and comprehension activities that previously required tedious human input, provides one tangible direction. This is an important instrument for language preservation and cultural archiving; yet studies linking the use of AI and language maintenance have not been well documented, perhaps because many studies of AI are driven from cultural settings not facing language death. Yet, consideration is given to preserving information published on the Web, or digitally mediated through printed publications then preserved as an “online” version after-the-fact, usually in English and detailed with an emphasis that the information is unique, valuable or, likely, useful for future data-mining. Hence, much of such practice is situated in Western nations that have the luxury, and, importantly, technical resources, for such an undertaking (Costa, Gomes, & Silva, 2017).

Of course, one vulnerability is that consensus seems hard to reach on what should or shouldn't be preserved. Of relevance might be the neo-Marxist, or perhaps libertarian, idea that digital preservation thwarts the more nefarious hostaging of information by big-data companies, which store and make use of human data for profit, rather than for pro-human futures (Day et al., 2015). Ultimately, however, this suggests that preservation of heritage and data, independent from such companies, is necessary to ensure appropriate representation of a given event, for prosperity, then. Yet, the socio-technical commoditisation of our data, hence heritage data, is now an intrinsic feature of the contracts of encapsulated interest that define much of our digital

existence, and, meanwhile, Internet activity is often heavily influenced by politics beyond the West (Day & Skulsuthavong, 2021a). As Foucault (1966/1970, p. 183) remarks, in any "...given moment, there is always only one episteme that defines the conditions of possibility of all knowledge..."

Consequently, colonial legacies, both internal and external, seek to maximise the storage and security of Western data expressed digitally, but little resource, or even consideration, is given to those cultures for whom data preservation would be a necessity. Thus, the data-value and act of such preservation inevitably creates a point of paradox; first, consensus must be reached on whether something should be preserved, and second, the act of such preservation must be supported by the resources of an institution with sufficient capabilities. Often, these institutions are situated far from events that might cause language death; a gap emerges where a form of digital colonialism occurs: largely Western narratives dominate and decide the social and cultural fates of minorities. Such forces thus hold the power to decide, then, whether a language can be preserved, along with its culture. In this sense, then, because of the profound data implications of storing culture and language in a digital archive, we are inevitably reliant on the demands of large, socio-technically complex western institutions to determine what is, or isn't, worth retaining.

Such preservation is not assured to prevent language death, identity loss, or cultural erosion. However, this problem is not new; the management and manifestation of knowledge are often inseparable from debates on truth, as well as power, which are subjectively defined. As Foucault (1980, p. 109) remarks, "Each society has its regime of truth... that is, the types of discourse which it accepts and makes function as true...". Therefore, as technological developments advance our capacity to record, preserve, and protect the past, along with endangered languacultures, it cannot be assumed that institutions with power will seek to do so, without bias. Indeed, some might seek to write the narratives of "their truths", as has been seen historically in numerous instances, especially in more authoritarian places. Moreover, the logistic complexities faced by archival of large amounts of data are problematic even within the most privileged countries in the world. Various organisations, which include US institutions such as The Internet Archives and The Library of

Congress, seek to preserve what we might term a “pro-human” Web whilst challenging big-business monetisation of data archiving (Day et al., 2015). The Internet Archives does so using automated and AI-driven crawling methods to save billions of web captures that, in 2021, took up 45 petabytes (or 45,000,000 gigabytes) of storage space (Kramer, 2021).

This is a lot of data; modern challenges of ensuring the preservation of such storage increase, naturally, along with the amount of data, which is challenging for developing nations’ archival efforts. Colavazzi et al. (2021) argue there is no clear study of the relationship between AI and archiving, though they draw attention to studies which highlight use for AI in the digital history and cultural heritage sector. Naturally, machine learning and automation lend themselves well to document processing, record-keeping, and other forms of digital transference that suggest a forthcoming interdisciplinary field joining Computer Science and History. In a similar regard, we note that there is not yet a clearly defined body of literature linking AI to its use in preserving endangered languages and enabling heritage revitalisation. Hence, we present it a need for it here. Indeed, much of language-related AI discourse on the area of cultural preservation focuses on transcription and translation, so accessibility of popular languages through the development of multilingual archives built on linked open data, as well as the enrichment of tools to make sense of the scale and scope of such huge data-sets (Wilde & Hengchen, 2017). Less consideration has been directed towards how AI might be used to preserve endangered languages, when every few months a language is suggested to vanish, along with its last speaker. Admittedly, Microsoft now pushes AI development towards preserving languages. Much of their efforts rely on neural machine translation models and deep learning strategies. For example, in January 2021, in partnership with the Government of Nunavut, Microsoft took steps to enhance text translation for the Inuktitut dialect of Inuktitut and add the Inuinnaqtun dialect to the Microsoft Translator (Peesker, 2021). Listed on UNESCO’s list of endangered languages, Inuinnaqtun is now a focus of preservation, as community members are crowd-sourced to increase the longevity of the language. The study established how, within the same culture, different generational groups of an endangered language may not be able to communicate with each other, thereby improving cross-generational

communication capacity. This is important. As Nuwer (2014) emphasises:

“Over the past century alone, around 400 languages – about one every three months – have gone extinct, and most linguists estimate that 50% of the world’s remaining 6,500 languages will be gone by the end of this century...Today, the top ten languages in the world claim around half of the world’s population. Can language diversity be preserved, or are we on a path to becoming a monolingual species?”

In 2021, the Hellenic Ministry of Culture and Sports partnered with Microsoft and A_DA to create an AI mapped model of Ancient Olympia called *Ancient Olympia: Common Grounds*. A virtual augmented reality, 3D technologies and drones with high resolution cameras were used to document the monuments, which were then combined with archaeological data to create models in case of further site deterioration (Lu, 2021). These demonstrate, then, two brief examples of collaboration between actors to preserve and educate, made possible by AI and Web technologies (Day, 2019). To this end digital heritage, the Web and AI, towards the goal of preventing language death, and promoting identity preservation, form an important trinity, and research direction.

Conclusion

Sociolinguists have argued that language death brings socio-political and socio-ecological loss. The erosion of language, and the role of politically dominant languages, creates implications that are both hurtful and symbolically significant. We have argued for a view of languages as social-semiotic systems embodied in human (neuro-)cognitive apparatus. Language death, which is often rooted in (neo-)colonialism, therefore, reflects a stop in the evolution of cultural systems, biology, knowledge and expression. It is a trauma, which erases traditions, genealogy, and marginalises those languages, alongside knowledge systems, deemed undesirable by (neo-)colonialist forces. Within this article, therefore, we explored a small case study of Māori languaculture in Aotearoa (New Zealand). The case of Māori tells us the importance of two things to combat language erosion: anti-colonialism and adapting to current contexts, particularly via digital technologies and AI.

Our analysis bodes practical implications for language maintenance and revitalisation, concluding that sociolinguistic practitioners should consider a socio-cognitivist, as well as socio-technical, paradigm for language intervention. Additionally, we highlighted leveraging AI technologies towards languacultural preservation to limit the destructive impact of language death, offering two examples of a growing field. Within this field, however, we stressed the economies of scale, survival, and sustainability that impact cultural heritage and preservation of endangered languages: the tentative issues of using costly AI resources, to preserve. We offered a narrative that suggests heritage preservation, via AI and other digital technologies, entails an inevitable reliance on powerful, likely-Western institutions. Paradoxically, preserving endangered languages, often colonially eroded, currently requires Western-funded AI. We must, now, negotiate the priorities of such institutions, rather than just preserving data for encapsulated mining, especially when governments rely on big business and data politics, not always pro-human (Day & Skulsuthavong, 2019; 2021a).

Our article presents a step towards an AI-driven future for the preservation of languaculture. However, whilst we have sketched a foundation, we acknowledge that our presentation is limited by the lack of a concrete plan for languaculture preservation in the age of AI. We consider this a temporary problem. The digital age demands that sociolinguists consider languaculture in forms beyond what traditional Linguistics might delineate. For example, social media, itself, is a new form of expression, filled with different communication acts. These could, given the mercurial nature of the Internet, become endangered. As seen in Myanmar, Thailand, and across Asia during 2021 and 2022, governments have sought to minimise social media expression, to conceal human rights violations and violence. Internal colonialism, then, is also a threat to languaculture. The widening scope of instantaneous expression, on the Internet, fuels discourse, power, and truth (Day & Skulsuthavong, 2021b). Treating such complex issues, concerning personal and national identity, as matters of languaculture and sociocognitive transmission, then, requires specialised strategies tailored to local contexts. This is no easy task but is essential for disentangling (neo-)colonialism and advancing a more sustainable pro-human world, even when oppressive forces

would see languacultures erased not only from the digital record but people's minds and futures.

Acknowledgements

We acknowledge the Māori of Aotearoa (New Zealand), for their efforts to preserve.

References

- Agar, M. (1993). *Language shock: Understanding the culture of conversation*. William Morrow.
- Agar, M. (2019). *Culture: How to make it work in a world of hybrids*. Rowman & Littlefield.
- Barrett-Walker, T., Plank, M. J., Ka'ai-Mahuta, R., Hikuroa, D., & James, A. (2020). Kia kua te reo e rite ki te moa, ka ngaro: Do not let the language suffer the same fate as the moa. *Journal of the Royal Society Interface*, 17(162). <https://doi.org/10.1098/rsif.2019.0526>
- Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge University Press. (Original work published 1972).
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). Harvard University Press. (Original work published 1979).
- Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Polity Press. (Original work published 1982).
- Butler, C. S. (2013). Systemic functional linguistics, cognitive linguistics and psycholinguistics: Opportunities for dialogue. *Functions of Language*, 20(2), 185–218. <https://doi.org/10.1075/foL.20.2.03but>
- Chew, P. G. -L. (2013). *A sociolinguistic history of early identities in Singapore: From colonialism to nationalism*. Palgrave Macmillan.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.
- Chomsky, N. (1988). *Language and politics*. Black Rose Books.
- Costa, M., Gomes, D., & Silva, M. J. (2017). The evolution of web archiving.

- International Journal on Digital Libraries*, 18, 191–205.
<https://doi.org/10.1007/s00799-016-0171-9>.
- Crystal, D. (2002). *Language death*. Cambridge University Press.
- Day, M. J. (2019). *Teaching the web: Principles for web education* [Doctoral thesis, University of Southampton]. University of Southampton Institutional Repository.
- Day, M. J., & Skulsuthavong, M. (2019, November 6–9). *Web science in SE Asia: Cultivating a 'Thai digital renaissance' through (re)introducing an interdisciplinary science in higher education* [Conference session]. Proceedings of the 7th Annual ANPOR Conference: Power of Public Opinion and Multicultural Communication Toward Global Transformation, Chiang Mai, Thailand.
- Day, M. J., & Skulsuthavong, M. (2021a). Towards social transformation in Thailand: Orwellian power struggles and 'digital' human rights under the socio-technical Thai internet panopticon. In C. Yamahata, D. M. Seekins, & M. Takeda (Eds.), *Social transformations in India, Myanmar, and Thailand: Volume I*. Palgrave Macmillan.
https://doi.org/10.1007/978-981-15-9616-2_17
- Day, M. J., & Skulsuthavong, M. (2021b). Newton's socio-technical cradle? Web science, the weaponisation of social media, hashtag activism and Thailand's postcolonial pendulum. *JOMEC Journal*, 16(1), 100–129.
<https://doi.org/10.18573/jomec.207>
- Day, M. J., Carr, L., & Halford, S. (2015, June 28–July 1). *Developing the 'pro-human' web* [Conference session]. Proceedings of the 2015 ACM Web Science Conference, Oxford, UK.
<http://doi.org/10.1145/2786451.2786458>.
- de Saussure, F. (1959). Course in general linguistics (W. Baskin, Trans.). In C. Bally & A. Sechehaye (Eds.), *Philosophical Library*. (Original work published 1916).
- Derrida, J. (1982). Signature event context (A. Bass, Trans.). In J. Derrida (Ed.), *Margins of Philosophy* (pp. 307–330). The Harvester Press.
- Du Bois, W. E. B. (1899). *The Philadelphia Negro*. University of Pennsylvania Press.
- Dunlop, M. (2021, July 28). *University academics' claim mātauranga Māori 'not*
-

- science' sparks controversy.* Radio New Zealand.
<https://www.rnz.co.nz/news/te-manu-korihi/447898/university-academics-claim-matauranga-maori-not-science-sparks-controversy>
- Erikson, E. H. (1968). *Identity: Youth and crisis*. W.W. Norton & Company.
- Erikson, E. H. (1970). Autobiographic notes on the identity crisis. *Daedalus*, 99(4), 730–759. <https://www.jstor.org/stable/20023973>
- Feldman, J. (2006). *From molecule to metaphor: A neural theory of language*. MIT Press.
- Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. Routledge. (Original work published 1966).
- Foucault, M. (1980). *Power/Knowledge: Selected interviews & other writings 1972–1977*. Harvester.
- Foucault, M. (1995). *Discipline and punish: The birth of the prison*. Vintage Books.
- Geeraerts, D., Kristiansen, G., & Peirsman, Y. (Eds.) (2010). *Advances in cognitive sociolinguistics*. De Gruyter Mouton.
- Halliday, M. A. K. (1978). *Language as a social-semiotic: The social interpretations of language and meaning*. Arnold.
- Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Deakin University Press.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. Hodder Arnold.
- Joseph, J. E. (2018). *Language, mind and body*. Cambridge University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.
- Lamb, S. (1999). *Pathways of the brain: The neurocognitive basis of language*. John Benjamins Publishing.
- Lamb, S. (2020). *The field of consciousness as a network of relationships: Support for idealism from theoretical linguistics* [Video presentation]. TSC Consciousness Conference, Tucson, AZ, United States. <https://www.youtube.com/watch?v=icqiPChQjI>
- Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.

- Latour, B. (1993). *We have never been modern* (C. Porter, Trans.). Harvard University Press. (Original work published 1991)
- Leibowitz, B. (2017). Cognitive justice and the higher education curriculum. *Journal of Education*, (68), 93–112. http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S2520-98682017000100006&lng=en&tIng=en
- Low, D. S., Aung, M. T., & Day, M. J. (2020). Cognitive sociology: developing the 'Diversity Pathways' model in cultural neuroscience. *Human Behavior, Development and Society*, 21(4), 66–77. <https://so01.tci-thaijo.org/index.php/hbds/article/view/243179>
- Lu, F. (2021, November 24). Ancient Olympia is digitally revived in Microsoft's mixed reality exhibition. *Jing Culture & Commerce*. <https://jingculturecommerce.com/ancient-olympia-common-grounds-microsoft-ai-ar/>
- Malabou, C. (2012). *The new wounded: From neurosis to brain damage* (S. Miller, Trans.), Fordham University Press.
- Matthiessen, C. (2012). Systemic functional linguistics as applicable linguistics: social accountability and critical approaches. *DELTA Documentação de Estudos em Lingüística Teórica e Aplicada*, 28(spe), 437–471. <https://doi.org/10.1590/S0102-44502012000300002>.
- Mcneill, I. (2021, July 12–13). *Internet freedom and digital authoritarianism during the Covid-19 pandemic: The need for a right to the internet* [Conference session]. International Virtual Conference on Emerging Research in Politics, Development, and Human Rights, Bangkok, Thailand.
- Ministry for Ethnic Communities. (2021, July 9). *Languages spoken in New Zealand*. <https://www.ethniccommunities.govt.nz/resources-2/our-languages-o-tatou-reo/languages-in-new-zealand/>
- Morrison, S. (2019). Kotahitanga in te reo Māori revitalisation. *Kura nui o Waipareira*, 3, 8–13. https://wairesearch.waipareira.com/dz_research_journal/kotahitanga-ma-te-kotahitanga-e-whai-kaha-ai-tatou-in-unity-we-find-strength/
- Nuwer, R. (2014, June 6). Languages: Why we must save dying tongues. *BBC Future*. <https://www.bbc.com/future/article/20140606-why-we-must->
-

save-dying-languages.

- O'Toole, M. (2020). Responsibility, language movement, and social transformation: The shifting value of te reo for non-Māori in Aotearoa New Zealand. In L. Siragusa & J. K. Ferguson (Eds.), *Responsibility and language practices in place* (pp. 195–212). Finnish Literature Society.
- Peesker, K. (2022, February 1). Government of Nunavut preserving endangered Inuit languages and culture with the help of artificial intelligence and Microsoft. *Microsoft News Centre*. <https://news.microsoft.com/en-ca/2022/02/01/government-of-nunavut-preserving-endangered-inuit-languages-and-culture-with-the-help-of-artificial-intelligence-and-microsoft/>.
- Radio New Zealand. (2019, July 11). Suicide rate among Māori men rises to highest in decade. *Radio New Zealand*. <https://www.rnz.co.nz/news/national/394108/suicide-rate-among-maori-men-rises-to-highest-in-decade>
- Reedy, T. (2000). Te reo Māori: The past 20 years and looking forward. *Oceanic Linguistics*, 39(1), 157–169. <https://www.doi.org/3623221>
- Salö, L. (2019). Thinking about language with Bourdieu: Pointers for social theory in the language sciences. *Sociolinguistic Studies*, 12(3–4), <https://doi.org/10.1558/sols.32916>
- Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Newbury House.
- Smith, L. T. (2012). *Decolonising methodologies: Research and indigenous peoples*. Zed Books.
- Tau, T. M. (2001). The death of knowledge: Ghosts on the plains. *New Zealand Journal of History*, 35(2), 131–152. http://www.nzjh.auckland.ac.nz/docs/2001/NZJH_35_2_01.pdf
- Te Puni Kōkiri. (2019). Pūrongo-ā-tau | Annual report for the year ended 2019. *Ministry of Māori Development*. <https://www.tpk.govt.nz/en/a-matou-mohiotanga/corporate-documents/purongoatau-annual-report-2019>
- Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. MIT Press.
- Wilde, M., & Hengchen, S. (2017). Semantic enrichment of a multilingual archive with linked open data. *Digital Humanities Quarterly*, 11(4). <http://www.digitalhumanities.org/dhq/vol/11/4/000328/000328.html>

Dylan Scott Low

Nacionalinis Taivano normalusis universitetas, Taivanas
dylanscottlow@outlook.com

Isaac Mcneill

Mahidolio universitetas, Tailandas
isaac.mcn@student.mahidol.edu

Michael James Day

Xi'an Jiaotong-Liverpool universitetas, Kinija
drmjday@outlook.com

**NYKSTANČIOS KALBOS: SOCIALINIS IR KOGNITYVINIS
POŽIŪRIS Į KALBŲ MIRTĮ, TAPATYBĖS PRARADIMĄ IR
IŠSAUGOJIMĄ DIRBTINIO INTELEKTO AMŽIUJE**

Santrauka. Sociolingvistai teigia, kad kalbų mirtis reiškia didelę kultūrinę, asmeninę ir ekologinę netektį. Sociokultūriniai ir sociopolitiniai veiksniai didina kalbos eroziją ir skatina jos išstūmimą kita, labiau dominuojančia kalba. Taigi klausiamo: kokie sociokognityviniai principai daro kalbos mirtį skaudžia ir simboline? Šiame straipsnyje pateikiame sociokognityvinį kalbos mirties aiškinimą, greta kognityvinės lingvistikos požiūrio pateikdami Halliday'aus požiūrį į kalbą kaip į „socialinę-semiotinę“ sistemą. Toliau kontekstualizuojame kalbą kaip neatsiejamą nuo kultūros, remdamiesi Bourdieu sociologine mintimi. Teigiame, kad kalbos mirtis sukelia psichologinę traumą, nes sunaikinama kultūrinė genealogija ir prarandamos žinios, susijusios su asmeniniu savęs įsivaizdavimu ir tapatybe. Šiuo tikslu pateikiame maorių kalbos kultūros Aotearoa (Naujoji Zelandija) atvejį, kuriame atsekame kolonializmo ir marginalizacijos poveikį iki dabartinių pastangų ir siekių užtikrinti maorių kalbos kultūrinį išlikimą ir susigrąžinti vietą Aotearoa kaip gerbiamai žinių sistemai ir išraiškos priemonei, ypač sociotechninėje dirbtinio intelekto ir interneto eroje. Teigiame, kad mūsų analizė turi praktinės reikšmės kalbos išlaikymui ir atgaivinimui, ir darome išvadą, kad praktikuojantys sociolingvistai turėtų apsvarstyti sociokognityvistinę ir sociotechninę kalbos intervencijos paradigmą. Pabaigoje aptariame dirbtinio intelekto technologijų panaudojimą kalbos paveldui, archyvavimui ir išsaugojimui, siekiant sumažinti destruktivų kalbos mirties poveikį.

Pagrindinės sąvokos: kalbos mirtis; kultūra; tapatybė; pažinimas; sociolingvistika; dirbtinis intelektas.