Reflections on documenting the lexicon

Keren Rice
University of Toronto

The lexicon presents unique challenges in language documentation. This reflection reviews some of those challenges, focusing on two major areas, what I have learned over time about what is important to document and the creation of dictionaries. Throughout I stress the value of considering the lexicon broadly, and, in the situation that linguists are involved, of working closely with speakers and community members in all stages of decision making, from what to document to how to spell, to how to represent meanings. N. Scott Momaday writes of words as medicine, and this is important to keep in mind in lexical documentation—one is engaging with worldview. The responsibility then of documenting the lexicon is large, and the stakes are high, given how words give deep insight into ways of being.

1. Introduction[1] The lexicon presents a challenge in language documentation, as Haviland (2006: 129) writes in his seminal paper on documenting lexical knowledge: "In the Boasian trilogy for language description of grammar, wordlist, and text, it is surely the dictionary whose compilation is most daunting. The process begins with a learner's first encounters with a language, and it ends, seemingly, never. Worse, it is an endeavor fraught with doubt, centrally about when enough is enough both for the whole - ...- but also for any single putative dictionary entry, given the apparent endless variety of nuance and scope for words and forms, not to mention the idiosyncrasies of compounds or derived expressions..." Frawley, Hill, and Munro (2002: 1) write of how a dictionary project goes on and on, "expanding from a modest list of words and glosses to something like a cultural encyclopedia." In this reflection, I identify some of the challenges in documenting the lexicon. My choices are idiosyncratic, reflecting my interests, but are, I suspect, of broader interest as well. In what follows I first reflect on documenting the lexicon and then on the generally expected output of documentation of the lexicon, dictionaries.

---

2. Documenting the lexicon   A useful starting point in addressing lexical documentation is the description of communicative events laid out in Himmelmann (1998). He identifies a cline of event types based on planning, ranging from unplanned to planned: exclamative (ouch!, fire!), directive ('scalpel', greetings, small talk), conversational (chat, discussion, interview), monological (narrative, description, speech, formal address), and ritual (litany). While Himmelmann has little to say directly about the lexicon, the very division into different event types suggests the challenges of documenting the lexicon—there is certainly vocabulary that cross-cuts event types, but there are lexical items that are likely to occur in one type and not in others. Thus, in order to obtain a lexicon that is both broad and deep, a considerable corpus must be developed.

In an ideal world, documentation would cover all the communicative event types, from exclamatives to ritual speech, and other uses of speech that might be found. Lexical documentation would include vocabulary gleaned from different communicative event types with different participants, and from a range of semantic domains, approaching, as Frawley, Hill, and Munro (2002:1) write, a cultural encyclopedia. All material would be audio recorded, with much video recorded, and then material would be drawn from those recordings, supplemented with additional material gathered through other methods, as needed.

This is, of course, the work of lifetimes. What should be prioritized? What can wait? There are no simple, straightforward answers to these questions but I will reflect on some of what I have learned from involvement with such work over some time period.

2.1 Documenting the 'unusual'   I begin at a level of speech that I did not consider when I began to do fieldwork, what I will call interjections. I have a grandchild who is being raised bilingually, in English and French, with more exposure to French than to English. What does he take to in English? He picks up on words like 'wow', 'oopsy-daisy', 'oops', 'uh-oh', 'ouch', and so on very quickly, and he likes to use them, and to reflect on the difference between words like 'oops' and 'uh-oh.' I have not heard him using this type of French word in English where an English one is appropriate, although he frequently uses French nouns and verbs when speaking with monolingual English speakers. There is something special about this part of the lexicon, and, while I had not thought to document it, although I heard it, watching my grandson acquiring two languages makes me realize how important such vocabulary is—it gives some kind 'feel' to the language. Similar is baby talk, talking to pets and animals, sounds that animals make, colloquial expressions, directives, and other communicative types where words or set phrases are likely an appropriate unit. Meanings might be difficult to express in another language, but this type of vocabulary is special, and knowledge of such words can be taken as a sign of cultural knowledge.

2.2 Etymologies   My next point is something that I have come to understand from working with Elders from several Canadian communities over the years, namely trying to understand what some think of as the 'true meaning' of a word. Some question the value of studying etymology. Mosel (2011:349), for instance, in writing about lexicography, suggests that although many people are interested in the history of languages, documentation of etymology should be postponed, with documentation of the living language taking priority. Whaley (2011:343) asks if thinking about etymology might be a case of "loving the language more than loving its speakers." Kroskrity (2015:151) notes a conflict in that seeking to understand the etymology of words can be

interpreted as privileging the past rather than the present. Nevertheless, in my experience an understanding of what words mean, where they come from, and their internal parts is valued by many involved in understanding what the values of their society are, and seeking to rebuild strengths that have been challenged through colonization. This is not to privilege the past, but to dream for the future. Perhaps the people who are keen about this are the philosophers and historians of their societies, but their desire to understand those values is important to them. For instance, Goulet and Goulet (2014:60), in their book on Nehinuw concepts and Indigenous pedagogies, discuss life force, writing that Nehinuw illustrates the interactive, dynamic process of causal forces that is deeply embedded in Nehinuw traditional culture. As one example, they give the root/stem waso 'she, he, it shines,' noting that it occurs in words like 'sun' and 'stars,' and also in the word awasis 'child,' or 'the little being that shines,' with children epitomizing the "light, sparkle, and vibrancy of life" in traditional Nehinuw culture. Drabek (2018) writes of how understanding the meanings of words of her ancestral language Kodiak Alutiiq help her to see the world in a different way. She considers the word, -imaq 'sea, ocean,' also used for 'a liquid contained inside' and contents. She writes how this word inspires her—*imartuq* 'it is full,' *imaituq* 'it is empty,' *imasuugtua* 'I feel depressed, or sad, I am down-hearted, I have a sinking feeling of foreboding,' or, more literally, 'I am searching for my contents.' Kroskrity (2015:151), focusing on designing a dictionary, discusses a Tewa verb that gives a sense of scarcity, at least from an historical perspective. He notes that the verb is restricted in the objects it can refer to, with those objects being things that were precious or vital to the well-being of the community. While much has changed, understanding the lexical semantics of this verb gives clues into values and worldview. Thus, understanding etymologies and lexical semantics can lead to insights that take one far beyond the language, to understanding values. This is, for some, not something to be disregarded as the etymology and meaning provide links between past, present, and, hopefully, future.

2.3 New vocabulary Another controversial area is new vocabulary, both loanwords and newly created vocabulary. Loanwords may or may not be recognized as such by speakers, depending on how much they are integrated into the language, broader knowledge of the speaker, and so on. New words are created to represent new things. Are such words included in documentation? The answer could be yes or no—although the 'ideal' documentation project might include new words as well as information about the process by which the language captures 'new' ideas, whether those new words are included in a lexicon depends on what is privileged in the particular context. It would be difficult to document without encountering loanwords and new words if speakers talk about a wide variety of topics, different communicative events with different participants are recorded, and so on. But if the focus is on tradition, such words might not occur.

The Alberta Elders' Cree Dictionary, compiled by Cree elders (LeClaire and Cardinal 1998), comments on new words: "A dictionary of this sort is not just a collection of words and their meanings, but represents something of what the community it serves requires. Hence, we have incorporated suggestions from a wide variety of Alberta bands for making this dictionary more usable for their members. .... [W]hat words we thought helpful, though not yet accepted widely by Cree speakers, or words that reflect recent English influence, or idiomatic Cree that did not appear connected directly to traditional Cree usages are ... found in the supplemental ...." The Elders who compiled the dictionary

believed that an indication of vitality of a language was the ability of speakers to talk about what was around them.

2.4 Documentary methods   It is worthwhile to very briefly consider methods of documenting the lexicon. Himmelmann's focus on communicative events can present challenges in documenting the lexicon, as much might not emerge in even a reasonably large corpus. In documenting semantic domains, there has thus been continuing emphasis on ways of documenting the lexicon that value teamwork with local and academic experts; see Evans (2012) for an anecdote on the importance of this. Many people I have worked with are keen to work on vocabulary in particular semantic domains, not just those that might be considered of relevance from a cultural perspective, but also things like types of footwear (including, in addition to words for moccasins, mukluks, and the like that are considered culturally significant, also words for running shoes, high heels, and so on). As one seeks to document the lexicon as fully as possible, awareness not just of the past, but also of the present, matters.

3. Thinking about dictionaries   Ogilvie (2011:402), examining the effects of language documentation on lexicography, concludes that "the lexicographer cannot ignore the new focus on primary data; the new recognition of the importance of collaboration and involvement of the speech community in the dictionary-making process; the new concerns for accountability and ethics; the new concern for storage and accessibility of archived dictionary materials; and the new possibilities that technology brings to both the content of dictionaries and their compilation." Ogilvie writes (2011:393) that dictionaries "have begun to blur the boundaries between documentation and description," commenting on how dictionaries are repositories for primary data, including images, sound, and video. While this blurring exists, I reflect on documentation—collection of primary materials— and description in the form of dictionaries separately, now addressing dictionaries.

Perhaps the major change since Himmelmann (1998) involves developments in technology. These allow for more kinds of dictionaries: talking dictionaries, dictionaries with videos, user-driven dictionaries. I do not pursue technology but would be remiss in not noting its tremendous role in advancing lexicography. Nor do I address the many other issues that arise in the discussion of dictionaries, including the nature of the word, the challenges of lexical analysis, and the content and organization of dictionaries.

In documentation, dictionaries are often designed with language sustainability or reclamation in mind (this is not to devalue fuller dictionaries, but such dictionaries take years and having products along the way is of value). While defining goals is worthwhile in determining structure and content, goals alone do not provide easy answers for many questions. I address three, standardization, meanings, and, in a different vein, issues of control.

3.1 Standardization   When I first began working with people on a dictionary, spelling seemed like a relatively unimportant issue. At that time, several years had been spent by committees working on standardization of symbols and spellings, and decisions had been reached on symbols to use, on how to spell, and on principles to follow in light of variation by and between speakers.

Starting with symbols, I became unsure why IPA vowel values have been so widely adopted in situations where speakers are familiar with values associated with English

vowels. I have been asked numerous times to put something in a spelling that English speakers can relate to more easily (e.g., instead of tu, write too, with the vowel like that in 'too'). It would have been difficult for the standardization committees to change symbols, given the material that existed with spelling based on IPA vowels, but the initial decision to use IPA was perhaps ill-conceived, not taking into account the likely users of the dictionaries and the background that they had.

Work on a dictionary gave rise to other surprises, as some of the principles of standardization that had been agreed on by standardization committees became sources of concern. Rice and Saxon (2002) discuss this, as do Hinton and Weigel (2002), Mosel (2011:341), and others. Mosel suggests that while standardization is often a political matter that can be difficult to resolve, having a standardized spelling is nevertheless important in a dictionary. The standardization committees recommended a single spelling reflecting conservative usage. But I learned that community developers of a dictionary do not necessarily agree. As some I've worked with have said, spelling standardization might be conceived of as a western ideal that is not held in all societies. Standardized spelling might emerge over time, but need not be the starting point, as it excludes rather than includes people, privileges some over others, and makes the relationship between the oral and the written more obscure, something that was not valued—people wanted to hear the voice of the person as they read their words.

Standardized spelling has proven to be untenable in current dictionaries that grow out of community work. In a Dëne Sǫłiné Yatié dictionary (Kaulback, Catholique, Drygeese 2014:11), for instance, the editors write of varieties and choices about spelling: "The changes and this variability are problematic when one begins the process of recording these words and preserving them in a written form…. We are cognizant of the fact that there are speakers that use the K-dialect in the community and others who don't. We are also aware that some words have been shortened but there remains a longer—some would say purer—form of the same word still in use among some speakers. In an effort to create a resource that best represents the language of the community and accounts for this variability, we have included alternate spellings and pronunciations for many words …. We encourage the reader to find their language within these words recognizing, of course, that not all forms of a word may be accounted for. … The elders recognized the value of this dictionary and the integrity of the process, and contributed to it with all their hearts. …"

Variation, I have learned, is well accepted in many places, and singling out one variety as 'better' may well be culturally inappropriate, at least at early stages.

3.2 Meanings and cultural concepts   Much has been written on semantic fields such as traditional tools, kinship, toponyms, astronomy, cooking, ethnobotany, and government. These are important topics, and worthy of inclusion in a dictionary, and are often located in thematic dictionaries or thematic parts of a comprehensive dictionary. Semantic fields can bring surprises. Once I was working with someone on classifying words into semantic fields, and we were discussing what field the words 'bow' and 'arrow' belonged in. I assumed hunting, and was taken by surprise when the response was that they were toys. For some they would have been considered hunting, but they are no longer used for that today, and this was what mattered to the person who I was working with.

In addition to semantic fields, aspects of meaning exist that are difficult to capture in translation, but give clues into values and worldview. I illustrate with an anecdote. I did a lot of knitting when I was doing fieldwork, and people often commented to me

that I was wasting my yarn. I found this an odd, even offensive, remark—I didn't think that I was wasting it, but rather that I was using or transforming it. It took me time to understand what was going on, and it is easiest to talk about this by giving some cultural background. Rushforth and Chisholm (1991) discuss cultural persistence in the Sahtú people of Canada's Northwest Territories, where I was doing fieldwork, introducing a Sahtú concept of *séodįt'é*, or what a culturally ideal person is like. Such a person is described as shy, humble, non-imposing, careful, caring, reasonable, reserved, controlled, polite, industrious, generous, restrained. This concept is important for understanding the nature of the lexicon. There are numerous pairs of verb stems, both of which translate roughly the same. For instance, the stems *-ta* and *-ʔe* both mean do something with the foot. There are, however, subtle differences between these that are hard to translate into English, and these differences relate to *séodįt'é*. Returning to the anecdote, I learned that what was translated into English as 'waste' is not negatively valued in the language, while it is in English—to put this another way, *séodįt'é* is positive but the reverse is not negative. How to address this in a dictionary—I still don't know the answer—words like 'gentle' vs. 'rough', 'slowly' vs. 'quickly' are used but they do not capture the essence of the difference. I do know that understanding the difference between these words provides a clue to understanding values of deep cultural importance, something that is expressed many aspects of the language as well as in ways of living.

3.3 Control I end with a discussion of control. While this does not have to do with the lexicon per se, it is an important topic as societies grapple with issues of privilege and power. The lexicon is a topic that, in a project involving an outside linguist, requires close collaboration with speakers. In working on dictionaries in documentation, work that blurs the lines between academic and community is increasingly understood to be important in many settings. A dictionary is, ultimately, a product designed for use by a community. There probably is no one right starting point for a dictionary beyond what people are interested in. It might be that the interest is in cultural traditions. But it might be that it is in naming new things. A dictionary of the type people are familiar with for languages like English, French, and Spanish may be a vision, but may be a barrier to dictionary creation. In the end, dictionaries are lists of words and phrases organized in some way so that words can be found, with, minimally, information about pronunciation, meaning, and use. Starting with expectations about what a dictionary should be can produce dictionaries that are rejected by a community; working collaboratively, trying to understand the vision for a dictionary at a particular time, trying to let go of pre-conceived notions of what a dictionary must be can lead to a very different type of dictionary than one could have imagined at the start.

Words matter. The Kiowa novelist, poet, and essayist N. Scott Momaday (1968:89) makes this clear in the following quote: "Words were medicine; they were magic and invisible. They came from nothing into sound and meaning. They were beyond price; they could neither be bought or sold ..." As a reviewer remarks, in engaging with the lexicon, one is engaging with worldview and ways of being. Thus, the stakes of lexical documentation and the resulting products can be very high, and all engaged must consider seriously the import of what they do, while at the same time delighting in how much they learn, both expected and unexpected.

References, plus some works that provided inspiration

Drabek, Alisha. 2018. Echoes from my Kodiak Alutiiq ancestors. (`https://www.humansandnature.org/echoes-from-my-kodiak-alutiiq-ancestors`) (Accessed 25 April 2018)

Evans, Nicholas. 2012. Anything can happen: The verb lexicon and interdisciplinary fieldwork. In Nicholas Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*, 183–208. Oxford: Oxford University Press.

Frawley, William, Kenneth C. Hill & Pamela Munro. 2002. Making a dictionary: Ten issues. In William Frawley, Kenneth C. Hill & Pamela Munro (eds). *Making dictionaries: Preserving Indigenous languages of the Americas*, 1–22. Berkeley: University of California Press.

Genee, Inge & Marie-Odile Junker. 2018. The Blackfoot Language Resources and Digital Dictionary project: Creating integrated web resources for language documentation and revitalization. *Language Documentation & Conservation* 12. 274–314.

Goulet, Linda M. & Keith Goulet. 2014. *Teaching each other: Nehinuw concepts and Indigenous pedagogies.* Vancouver: UBC Press.

Grenoble, Lenore & Simone S. Whitecloud. 2014. Conflicting goals, ideologies and beliefs in the field. In Peter K. Austin & Julia Sallabank (eds), *Beliefs and ideologies in language endangerment, documentation and revitalisation. Proceedings of the British Academy 199*, 339–356. Oxford: Oxford University Press.

Haviland, John. 2006. Documenting lexical knowledge. In Gippert, Jost, Nikolaus Himmelmann & Ulrike Mosel (eds), *Essentials of language documentation*, 129–161. Berlin: Mouton de Gruyter.

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.

Hinton, Leanne & William Weigel. 2002. A dictionary for whom? Tensions between academic and nonacademic functions of bilingual dictionaries. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.) *Making dictionaries: Preserving Indigenous languages of the Americas*, 155–170. Berkeley: University of California Press.

Kaulback, Brent, Bertha Catholique & Dennis Drygeese (eds). 2014. *Dëné Sǫ́łiné Yatié. Ɂerehtł'ís. łuskëlk'e t'iné Yatié. Chipewyan Dictionary.* South Slave Divisional Education Council.

Kroskrity, Paul. 2015. Designing a dictionary for an endangered language community: Lexicographical deliberations, language ideological clarifications. *Language Documentation & Conservation* 9. 140–157.

LeClair, Nancy & George Cardinal, edited by Earl Waugh. 1998. *Alberta Elders' Cree dictionary / Alperta ohci kehtehayak nehiyaw otwestamâkewasinahikan.* Edmonton: University of Alberta Press.

Momaday, N. Scott. 1968. *House made of dawn.* New York: Harper & Row.

Mosel, Ulrike. 2011. Lexicography in endangered language communities. In Peter K. Austin & Julia Sallabank (eds.), *Cambridge handbook of endangered languages*, 337–353. Cambridge: Cambridge University Press.

Ogilvie, Sarah. 2011. Linguistics, lexicography, and the revitalization of endangered languages. *International Journal of Lexicography* 24(4). 389–404.

Rice, Keren & Leslie Saxon. 2002. Issues of standardization and community in Aboriginal language lexicography. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries: Preserving Indigenous languages of the Americas*, 125–154. Berkeley: University of California Press.

Rushforth, Scott & James Chisholm. 1991. *Cultural persistence: continuity in meaning and moral responsibility among the Bearlake Athapaskans.* Tucson: University of Arizona Press.

Whaley, Lindsay. 2011. Some ways to endanger an endangered language project. *Language and Education* 25(4). 339–348.

Keren Rice
rice@chass.utoronto.ca
orcid.org/0000-0002-8112-8908