> **Lexicography and Endangered Languages: What Can Europe Learn from the Rest of the World?**[^1]

SARAH OGILVIE

1 Introduction

In 1926, at the age of 24, the controversial American anthropologist Margaret Mead (1901-1978) was on her first field trip to Samoa. She wrote back to her PhD supervisor at Columbia University, the famous Professor Franz Boas (1858-1942), saying 'Through it all, I have no idea whether I am doing the right thing or not, or how valuable my results will be. It all weighs rather heavily on my mind'.[1]

This, I would like to suggest, is the sentiment of every fieldworker who is documenting a language for the first time, and, if the language is endangered, most probably for the last time. Linguistic documentation and description has traditionally entailed recording the language, transcribing the language, and writing a grammar of the language. Writing a dictionary of the language was more a stepping stone towards the grammar, rather than a goal in itself. Most dictionaries of endangered languages, therefore, are compiled by linguists or anthropologists who are not lexicographers. They learn the craft 'on the job', and most of these 'new lexicographers' – and I say this from personal experience – feel the same bewilderment as Margaret Mead: they have no idea whether they are doing the right thing or not, or how valuable their results will be.

This was certainly the case for me when, twenty years ago, I arrived in an Aboriginal community on the tip of Cape York Peninsula in remote northern Australia with a tape recorder, ten cassettes, a notebook and pencil, and a change of shorts and t-shirt. My task was to write a grammar and dictionary of an endangered Aboriginal language called Morrobalama. The language had never been recorded or written down before, and my task was to describe the language before the last two speakers died.

My only experience of dictionary writing at that time had been as pronunciation editor on the *Australian Concise Oxford Dictionary* 8th ed., responsible for creating the first Oxford dictionary with Australian rather than British pronunciations. That job did little to prepare me for

....

1   Letter from Margaret Mead to Franz Boas, 16 Jan, 1926. 'The Correspondence Between Margaret Mead and Franz Boas Exchanged During Mead's 1925-26 Samoan Research Project (and related material)' <http://sociology.uwo.ca/mead/>

writing a dictionary of a language that 1) I had never heard before, 2) had never been written down before, and 3) was spoken by two last speakers who were rarely sober enough to teach me their language. The safe and comfortable offices of Oxford University Press were in stark contrast to life in a community where I slept on the ground each night on a burnt-out mattress, ate fresh-water turtle, and generally felt like the 'stranger' on every level: culturally, linguistically, and socially.

When one crosses the boundary between one's own language and culture into another, one can't help but be changed by it. Claire Bowern calls this a 'peculiar displacement' in which 'the fieldworker is displaced from their own community and culture, and is sent to think analytically about another social and linguistic system'.[2] Daniel Everett described his experience of documenting the Pirahã language in the Amazon as akin to 'becoming an alien'.[3] He warned other field workers:

> You could become a 'freak' instead of an attractive person; an incompetent, instead of a respected professional; ugly instead of lovely; fat instead of average; stinky instead of normal-smelling; and on and on. You may go from being articulate and witty in conversation to being perceived as an infantile dullard who can barely function in conversation. You will go from having many friends to having none. From enjoying good company, to stark loneliness.

Twenty years ago, field lexicography was lagging behind commercial lexicography on all levels, and my experience of dictionary making in the field as opposed to the office certainly provided a stark contrast to my future experiences in lexicography, as I went on to be Senior Editor on the *Macquarie Dictionary* and various Oxford Dictionaries in Australia, and more recently on the *Shorter Oxford Dictionary* and the *Oxford English Dictionary* (*OED*) in the United Kingdom. In the world of language description, there was barely any overlap between field linguists and commercial lexicographers, and I found myself in the unusual position of combining the two. In recent years however, linguists have started to do innovative work on collecting primary data and rethinking the principles, theories, and practice of documenting languages and cultures. Their concern not only for language preservation but also for its maintenance and revitalization has meant that field linguists have had to rethink how to write dictionaries.

....

2   Bowern (2007:10)
3   Everett (2006:6)

What can we – as members of Euralex and as writers of European mainstream dictionaries and dictionaries of minority languages – learn from dictionaries produced by documentation linguists? What is the potential relationship between linguists and lexicographers? When I was an editor on the *OED*, I was the only one of forty editors trained in linguistics. I remember being surprised by this, and one day raising it with the Chief Editor. He explained that he preferred it that way because linguists thought too much about things. It is true that thought can slow things down in lexicography... Seriously though, an historian or literary scholar can often contribute more to historical lexicography than the specialist in linguistics. However, there have been changes in the area of descriptive and documentary linguistics in the past decade that suggest that linguists might have something to teach lexicographers. How might we all share our expertise with one another? Can we make more of the relationship between lexicographers and linguists than we have in the past?

2   Lesser Used Non-state Languages

The theme of this paper is specifically dictionaries of endangered languages, but the theme of the 14th Euralex International Congress is dictionaries of 'lesser used or non-state languages'. While all endangered languages fall in this category, not all lesser used non-state languages are endangered, i.e. endangerment depends on the degree of language shift. Twenty years ago, Joshua Fishman identified eight steps toward reversing language shift.[4] The steps progressed from the ultimate goal of step 1 – making a language the language of national government – to the easiest goal of step 8 – reconstructing the language and designing language learning programmes. Where a language sat on this spectrum was considered a barometer of its chances of being saved and revitalized. Speakers of non-endangered languages that are lesser used and non-state probably take Fishman's step 8 as a given, and step 1 as a real desire and possibility. Speakers of endangered languages, on the other hand, may strive for step 8 and not even dream of the possibility of step 1.

But that was twenty years ago, and many linguists see things differently now. They follow the lead of Leanne Hinton who shifted the focus from the national to the domestic, from the ultimate goal of government recognition and sanctioned use to the realization that languages must first be spoken at home by children if they have a chance of being spoken

anywhere.[5] This change in scholarship has affected in fundamental ways the approach to linguistic description and the nature, focus, and quality of documentation and revitalization programmes. I would like to suggest that it has also changed the nature, focus, and quality of dictionaries of endangered languages in ways that all of us can learn from, regardless of whether our aims are to promote our language to national or domestic level.

The past ten years have seen the emergence of new lexicographic policies and practices around the world that can be characterized by an innovative exploitation of new technologies, predominant use of oral as well as written sources, incorporation of pedagogical materials, and collaborative involvement of members of the speech community. For these reasons, this paper will focus specifically on lexicography of endangered languages around the world.

3   Endangered Languages

There is no doubt that one of the most important issues facing humankind today is the rate at which our languages are dying. On present trends, the next century will see more than half of the world's 6800 languages become extinct, and most of these will disappear without being adequately recorded.[6] Current language distribution shows that 96% of the world's languages are spoken by only 4% of the world's population. David Crystal calculated in his book *Language Death* that one language dies on average every two weeks.[7] And, of course, more is lost than mere words. As vehicles for the transmission of unique cultural knowledge, local languages encode oral traditions that become threatened when elders die and livelihoods are disrupted. When a language disappears so does a culture and a speech community's unique way of seeing and ordering the world.

What kinds of languages are we talking about? Let me play for you a few words of Yurok, a North American Indian language, with six fluent speakers left. In this case, the language has been documented (and I will talk about that later) but it remains to be seen whether it will be revived successfully. If not, it will die out in the next decade. Likewise this video of one of the last ten speakers of Kayardild, a language of northern Australia. In twenty years time, unless the language is properly documented and

....

5   See Hinton (1997) and Hinton and Hale (2001).
6   Krauss (1992), Crystal (2002:19).
7   Crystal (2002:19)

revived no one will be speaking Kayardild and cultural connections such as these will be lost.

4   Language Documentation and Description

Unless the academic community works swiftly with indigenous communities and NGOs in collaborative and innovative ways, most of this expressive diversity will disappear without being adequately recorded or given a chance of conservation and revitalization. An important first step in slowing down or reversing the process of language death is to document the language in the form of a dictionary. Using innovative lexicographic policies, practices, and technologies, the lexicographer is able to produce dictionaries that are useful to both communities and scholars; dictionaries that not only describe and preserve an endangered language – as was the goal of linguists in the past – but also help in the processes of maintenance and revitalization.

Writing dictionaries of this kind is important on a number of levels. On an immediate level, as lexicographers, we have a duty to speakers of a language to record and describe their words with precision, accuracy, and in a way that is most useful to them. For those of us who are linguists, our linguistic theories depend on linguistic diversity and the rigorous description of that diversity. But more important, for humanity in general, is the need to preserve cultural diversity and knowledge systems that can be encoded in a dictionary.

For many years in descriptive linguistics, academics wrote dictionaries of endangered languages that were merely by-products of their primary aim – which was to describe the grammar of a language. My work on Morrobalama certainly fell in this category. But linguists and indigenous communities now recognize the important role that dictionaries can play in the documentation, preservation, and revitalization of endangered languages, and the past decade has seen linguists and anthropologists begin to focus on dictionaries as important tools and products in themselves.[8] These changes have been accompanied by new trends in Documentary Linguistics and Anthropology as priority research areas that deal with the principles, theories, and practice of documenting languages and cultures that are at risk.[9] In 1998, in a landmark article in the journal *Linguistics,* Nikolaus Himmelmann formally distinguished

8   See Frawley, Hill, and Munro (2002) for evidence of this.
9   See Himmelmann (2002), Woodbury (2003), and Austin (2006).

between language documentation and language description. The aims of language documentation were to record the primary data of language study, e.g. spoken and written texts which are transcribed, annotated with metadata, and archived for posterity. Language description, by contrast, was concerned with the secondary data of language study, e.g. analysis of primary data in the form of dictionaries and grammars. Since then, however, dictionaries of endangered languages have begun to blur the boundaries between documentation and description. More and more, they have become repositories for primary data which include images, sound, and video. This development has coincided with innovations in technology and documentation techniques thereby opening up the field of lexicography beyond academia so that academics are joined in the task by indigenous communities, educators, and certain NGOs whose work involves language support.

5 Compiling Dictionaries of Endangered Languages

For the endangered-language speech community, the most useful and relevant research outcome of field linguistics is usually the dictionary. Articles and books on syntax, morphology, or phonology have little relevance to indigenous speech communities. Dictionaries, on the other hand, are not only useful and functional texts, but emblems and tools of prestige which many communities use to boost their sense of identity and political profiles.

For the lexicographer, the field situation often presents a complex set of challenges that have an impact on lexicographic policies and practices. On top of the challenging living conditions, an undocumented language presents challenges relating to audience (are you writing for scholars or the speech community?), format (will it be a print dictionary, web-based, or electronic with imbedded pictures, sound, or video? Will the dictionary be linked to learning materials?), and compilation (what orthography and writing system will you devise? How do you list words in a dictionary if the language does not really have separate lemmas but rather joins up all the units of meaning into one polysynthetic word that we would probably call a sentence? How will the compilation involve the speech community? What software will you choose to accomplish this?). All of these issues – the audience, format and mode of compilation of the dictionary – will depend on region; health of the language and degrees of endangerment; community attitudes towards language, literacy, and learning; and access to electricity and internet.

The collaborative dictionary-making efforts of academics, community

members, and NGOs are producing dictionaries that are community-focused and collaborative in their compilation, content, and format. Currently, in response to different degrees of language endangerment, dictionary projects around the world fall into one of three categories: dictionaries for language preservation, dictionaries for language maintenance, or dictionaries for language revitalization. While this paper is not an exhaustive survey of projects around the world, I have chosen some examples of dictionary projects which have developed methodologies that nonetheless might have applicability to European dictionaries whether they be of minority or majority languages.

5.1 – Dictionaries for Language Preservation

In the Aslian (Mon-Khmer, Austroasiatic) languages of the equatorial forests of Malaysia, Niclas Burenhult is currently compiling dictionaries of Jahai, Semnam, Menriq, Batek, Lenoh, and Maniq. They focus on descriptions of unique ethnobiological knowledge about the forest and how to make a sustainable livelihood from it. In compiling the dictionaries, Burenhult faced tricky decisions relating to the order of entries, choosing not to order the headwords alphabetically but rather according to manner and place of articulation with left-to-right ordering rather than rhyming order, as is the tradition in many Austroasiatic dictionaries. At this stage, with no literate speakers, the dictionaries are primarily for preservation and scholarly purposes.

5.2 – Dictionaries for Language Maintenance

While access to computers and the internet is rare in many remote parts of the world, mobile phone access is not. In remote parts of Australia, for example, the presence of mining companies in the Outback has brought network access to areas that probably would not normally have been priority zones for telecommunication companies. Hence, perhaps surprisingly, people in remote Aboriginal communities currently own and use mobile phones more than any other form of technology. There has been a successful dictionary program by James McElvenny and Aidan Wilson at Sydney University, the Project for Free Electronic Dictionaries, to install dictionaries of endangered Australian Aboriginal languages on mobile phones.[10] Loaded on to a mobile phone via software called

10   The dictionary software for mobile phones can be downloaded at http://www.pfed.info/wksite

Wunderkammer, a Java ME MIDlet, each dictionary entry has a spoken pronunciation and many entries have pictures. Currently, the Wagiman language, spoken in the Northern Territory of Australia, is on mobile phones, and further projects are currently underway for Tuva, a language of the Ivory Coast, and Whitesands, a language of Vanuatu.

In recent years there has been a trend in endangered-language lexicography to produce small dictionaries of semantic fields. These are particularly suited to language maintenance, in the sense that breaking down the mammoth overall task of compiling a comprehensive dictionary into 'mini dictionaries', provides the speech community with quick access to a dictionary of their language for use in schools and the community in general. Ulrike Mosel and Ruth Spriggs compiled mini dictionaries of Teop, a language spoken in Papua, which covered semantic fields such as House Building, Body and Health, Fish, Shells, and Trees. The mini dictionaries were collaborative efforts with older speakers who assisted with editing, young speakers who checked the clarity of the entries, and children who gave feedback on the dictionary's lexical coverage (e.g. Teop children collected shells which they found missing in the first draft of the shell dictionary). Mosel and Spriggs found that collaborative lexicographic activities such as these promoted language awareness and pride in young speakers, the targeted demographic for successful language maintenance or revitalization. Being able frequently to present the speech community with tangible results of lexicographic work, in the form of mini dictionaries, rather than wait years for the completion of a comprehensive dictionary, has the additional benefit of demonstrating the lexicographer's commitment to language maintenance and revitalization in the community, and their ability to produce results.

5.3 – Dictionaries for Language Revitalization

It is in the area of language revitalization that the most exciting lexicographic work is taking place. Dictionaries written for revitalization have to address quite a complex set of issues relating to the stage of endangerment, level of literacy, and opportunity for capacity building and empowerment of community members to revitalize their language.

Dictionaries of all endangered languages have the added pressure of having to be compiled quickly, or at least the materials must be collected quickly, before the last speakers die. The Iquito Dictionary Project in northern Peruvian Amazonia, led by Christine Beier and Lev Michael, advocates a team-based and community-participatory approach to dictionary writing

which helps in fast collection of data.[11] The research team comprised 2-3 community linguists and 4-7 visiting linguists (professors and graduate students) who visited the field at the same time. The initial task of the visiting linguists is to help with capacity building and skills-transfer activities so that community members can be trained as 'community linguists', and work alongside the research team. In the case of Iquito, an Amazonian language with 25 speakers all of whom are over the age of 65 years, a few of the community members were immediately trained in basic aspects of descriptive linguistics and language documentation. Training of this sort is not always a straight forward process, as it is often the case that last speakers of endangered languages are not literate, and members of the community who are literate may not be proficient in any of the indigenous language. It is therefore important to incorporate literate adults as 'community linguists' and traditional speakers as 'language specialists'.[12]

The working schedule for team-based dictionary projects is highly structured. The collection of data for the dictionary takes the form of weekly data-gathering tasks or 'modules' – each task allocated to a different team member – the results of which are reported daily to the rest of the team in the form of a 'seminar'. In the case of the Iquito Dictionary project, this schedule of Module-and-Seminar continued every day during the two-month visits by the academics each summer for three years (2003-2006), and dictionary compilation continued throughout the year by the local community linguists. The language is now preserved in a printed bilingual *Iquito-Spanish Dictionary*, and the community linguists now teach Iquito language classes in the community's schools.[13]

Transfer of skills and capacity building are therefore responsible for turning what may have just been a language preservation dictionary project into a language revitalization dictionary project. The project trained a group of independent local experts – community linguists and language specialists – who could serve the community beyond the life of the dictionary compilation. The inclusion of graduate students in the research teams was also an ideal way of training and mentoring future lexicographers – all the while supporting their first experience of field lexicography with social, scholarly, and material infrastructure. Not only does this boost the numbers of linguists and anthropologists who learn

....

11   See Beier and Michael (2006)

12   See Beier (2009:4)

13   In addition to the dictionary, the IDLP (Iquito Language Documentation Project) team also produced grammatical analyses and an extensive collection of audio, video, and written texts which are described further in Beier (2009) and Michael (2009).

the art of lexicography in the field, but it also increases the productivity and amount of dictionary work carried out in any one field trip.

From the perspective of the visiting lexicographer, the team-based approach to dictionary-writing has the additional benefit of providing social support in what can otherwise be an isolating situation. However, the lexicographer must be careful that her/his integration into the speech community is not jeopardized by the comfort of socializing solely with other members of the visiting team. In an oral culture, the field lexicographer's access to words and language is increased by his/her ability to integrate into a speech community. Hence, within the team-based model of documentation the lexicographer must be careful not to rely too heavily on the social support of other visiting members of the team, especially if such socializing would neglect relationships within the community.

Another capacity building strategy in the rest of the world that supports dictionary making is the BOLD (Basic Oral Language Documentation) initiative in Papua New Guinea, in which Olympus has donated hundreds of voice recorders for traditional speakers to record their languages. This project, organized by Steven Bird, has a strict schedule of voice recording and transcribing, all of which can feed into dictionary-building. With about 850 languages, Papua New Guinea is the most densely populated region for language diversity in the world. The BOLD project provides Olympus VN5200PC digital voice recorders to one hundred speakers of different languages. Over a period of one year (February 2010 – February 2011), participants commit to a three-stage process: first, participants record 10 hours of culturally-rich speech (e.g. conversation, personal narratives, and idiomatic speech). The next stage involves re-playing the recordings and re-speaking the oral translation with another speaker on another digital recorder. This recording therefore contains not only the original recorded text but also a commentary on it. The third and final step involves choosing one or two segments of the original recording that amount to six minutes of spoken language, and transcribing it.

On average, BOLD participants spend one hour transcribing one minute of recorded text. Many of these languages have not been written down before, so the process of transcription in stage three will prompt the speakers and participants to think about the written representation of sound and the challenges of devising an orthography for their own languages.[14] These transcribed texts and primary data will go towards writing dictionaries

14   For more on native orthographies, see Harrison and Anderson (2006).

and grammars of the languages of Papua New Guinea, but, along the way, the process will have trained native speakers in the techniques of language documentation, created community interest and pride in their traditional languages, and, in many cases, prompted indigenous community members to think about their languages in a new way.

For critically endangered languages (those with no child speakers), it is not only necessary to record the language quickly, but it is important for the dictionary content to facilitate, or potentially facilitate, language revitalization. In addition to the resultant skills transfer from collaborative techniques of dictionary compilation, there are also mechanisms within the dictionary itself that can aid revitalization and make the text more appealing, functional, and useful to language learners, especially children. For example, for communities with computer and internet access, such as the Yurok North American Indian tribe in northern California, the dictionary entries can be linked to language memory tests and language learning exercises with dictionary audio files.[15]

Available free online, the Yurok Dictionary is similar in structure to the *Oxford English Dictionary* in that it is an historical dictionary which shows the use of Yurok words over time. It makes use of the fact that the Yurok language was recorded at different times throughout the twentieth century. Recordings of the language on wax cylinders have made it possible for Andrew Garrett at UC Berkeley to include a quotation paragraph after each definition showing how the word was used at different points of the twentieth century. For example, in 1902 and 1907, the language was recorded by the famous American anthropologist A. L. Kroeber (1876-1960); in 1927, it was recorded by the doyen of linguistics Edward Sapir (1884-1939); in 1958, it was recorded by the British linguist R. H. Robins (1921-2000); in 1980 and 1986 the same speaker whom Robins had recorded, Florence Shaughnessy, was recorded again by Paul Proulx and Jean Perry respectively; and finally in 2007, the last remaining speakers were also recorded. The Yurok Dictionary is able to supply each entry with recorded illustrative sentences from throughout the twentieth century (1902-2007), as exemplified at the entry *kwelekw,* adverb meaning 'well'. Illustrative sentences are linked to the larger texts in which they appear, and users see a picture of the original speaker and can read or listen to the original recordings of the entire stories, such as this recording of Mary Marshall telling Edward Sapir the story of 'Coyote Tries to Kill the Sun' in 1927 or Domingo of Weitchpec telling A. L. Kroeber the story of 'Buzzard's Medicine' in 1907.

....

15   http://linguistics.berkeley.edu/~yurok/web/random.html

There is one important difference between the historical examples in the Yurok dictionary as opposed to those in the *OED*. The Yurok examples are predominantly based on spoken, rather than written, evidence. Dictionaries of endangered languages are based on oral more than print culture which thereby captures more words from different genres. In my own work as a lexicographer in the UK, I had responsibility for non-European words in the *OED*, so I am aware of the restrictive implications of inclusion policies that require a minimum of five published citations over five years. This policy, based on concerns for the unreliability of spoken or unpublished sources, is particularly difficult to satisfy for words in English from parts of the world without established publishing traditions. A word in Philippine English may not appear five times in print but it may be used in the English of 42 million speakers. Hence I am aware of the hundreds of words that did not get in to the dictionary because of the bias in our European lexicographic tradition toward printed sources. Inclusion policies based on the number of citations from written sources get increasingly difficult to defend as technology improves our ability to capture, reproduce, and verify natural speech in natural contexts. Perhaps this is an area in which mainstream lexicography will follow innovations in field lexicography.

Unlike most lexicographers of minority languages in Europe, who are frequently native speakers of the language they are describing, lexicographers of endangered languages must undergo the slow process of learning the language they are describing. If they are writing dictionaries for language revitalization, they face the added challenge of not only learning the language themselves but also facilitating the learning (and teaching) of language for others within the community. In addition to creating a text – like the Yurok Dictionary – that facilitates language learning, the lexicographer may be in a position to empower native speakers and young adults in the community to work together so that young members acquire conversational proficiency in the traditional language. By doing this, the lexicographer can help to ensure that language learning becomes a part of the community culture beyond the life of the dictionary project. As explained by Chief Harry Wallace, the elected leader of the Unkechaug Nation (Long Island): 'When our children study their own language and culture, they perform better academically. They have a core foundation to rely on'.[16] The Africanist Paul Newman, however, criticizes these efforts by lexicographers and linguists because he argues that they should not

16   As quoted in 'Indian Tribes Go in Search of Their Lost Languages' *New York Times* 6 April 2010, C1.

become mere 'linguistic social workers' who waste their skills and time on the 'hopeless cause' of language revitalization.[17] Far from a hopeless cause however, there are numerous examples of lexicographers around the world who successfully negotiate a balance between dictionary work and revitalization work, and for dictionaries written with revitalization as one of the outcomes, many would argue that the two are inseparable. Indeed, many field lexicographers successfully facilitate language revitalization, and in turn these efforts result in increased dictionary use and ultimately a reinforcement of the lexicographer's raison d'etre.

One proven and successful methodology for bringing native speakers together with language learners is the Master-Apprentice Program, originally devised by Leanne Hinton, Nancy Richardson, and Mary Bates Abbott for revitalization of Californian languages.[18] By instituting this method while compiling the dictionary, the lexicographer lays the foundation for other one-on-one relationships between traditional speakers (the Masters) and language learners (the Apprentices). Hence, at the same time that the lexicographer learns the language from the Master, s/he also sets up a facility for language learning that can be replicated by other members of the community. The program advocates five main principles: 1) The Master and the lexicographer must not speak together in the dominant language (i.e. the language which is replacing the endangered language); 2) only oral (not written) language must be transmitted; 3) the lexicographer must be at least as active as the Master in deciding what is to be learned and in keeping communication going in the language; 4) learning must take place in real-life situations and traditional activities e.g. collecting food, going hunting, cooking, and doing crafts; 5) it must all be recorded or videoed for later analysis and use in the dictionary.

Advocating and practising a lexicographic methodology that facilitates the maintenance and revitalization of endangered languages is only part of the process. Ultimately, of course, whether or not a language survives – and the role that a dictionary plays in this process – will depend on the speakers themselves i.e. their attitudes towards the language in general and their willingness for inter-generational language transmission.

Activists for preservation of endangered languages often stress the urgency

....

17   Newman (1999) and Newman (2003).

18   See Hinton (1997) and Hinton (2001) for more information on the Master-Apprentice Program.

of capturing and saving languages before they disappear, arguing that it is literally a matter of life or death. Is it? The logical extreme of dictionaries for revitalization, of course, are those that are written from direct contact with no speakers at all. It is possible to revive a language from written sources alone (e.g. modern Hebrew) and every field lexicographer must hold in their mind the possibility that their own work may one day be used for such a purpose. In 1791, when the third President of the United States and the principal author of the Declaration of Independence, Thomas Jefferson (1743-1826), collected a wordlist from the last three speakers of Unkechaug, he probably had no idea that their descendents would be using his wordlist to revive the language on Long Island in 2010.[19]

Indeed, the current work by lexicographers of endangered languages will surely provide materials for language programs of the future. The exact sound, form, and structure of that language may not be exactly the same as that recorded by the lexicographer but the dictionary maker must be mindful of the possible future uses of her/his work. Unlike dictionary work on languages with established literary traditions, like those in Europe, the stakes are particularly high with endangered languages. The accuracy with which a lexicographer describes the sound, form, meaning, history, and usage of words from endangered languages may be the only lasting record of a language and culture, and future generations will depend on it in unforeseen ways: 'Would someone from 200 years ago think we had a funny accent?' asked Robert Hoberman, organizer of the Unkechaug revitalization, 'Yes. Would they understand it? I hope so.'[20]

Similarly, Natasha Warner and Quirina Luna are currently writing a dictionary of Mutsun, the language traditionally spoken south of San Francisco, California. It has been extinct, or 'dormant' as Warner and Luna prefer to describe it, since 1930, but the lexicographers are hoping that their dictionary will enable 'all interested members of the community to achieve reasonable fluency in (the revitalized form of) the language, at which point it is likely that some Mutsuns would be raising their children in Mutsun'.[21] The dictionary was compiled using original notes and materials by the early nineteenth-century Roman Catholic missionary, Felipe Arroyo de la Cuesta, and the early twentieth-century anthropologist, J. P. Harrington. In the 1920s, the eccentric Harrington

19   See 'Indian Tribes on Long Island Go in Search of Their Lost Languages' *New York Times* 6 April 2010 C5.
20   Robert Hoberman quoted in 'Indian Tribes on Long Island Go in Search of Their Lost Languages' NYT 6 April 2010 C5.
21   Warner, Butler, and Luna-Costillas (2006:259)

collected 36,000 pages of notes on Mutsun (within a two year period) from the last fluent speaker, an elderly Mrs Ascension Solorsano. These have been collated into a dictionary of headwords with a uniform orthography. The lexicographers were also faced with the task of inventing new Mutsun terms for the modern world, e.g. restaurant 'ammamsa' = eat+locative nominalizer.

The Mutsun dictionary initiative and the Unkechaug revitalization efforts both came out of a workshop organized by Leanne Hinton at UC Berkeley called 'Breath of Life'. Every two years, the Breath of Life workshop brings 60 people who identify as North American Indians to UC Berkeley for one week. They are united by one similarity: their traditional languages are extinct, but each person is accompanied by two mentors who are lexicographers or linguists. They spend the week receiving intensive training each morning in the basics of lexicography and linguistics. Each afternoon, they are shown how to use the rich linguistic and anthropological archives housed at UC Berkeley, and each evening the participants work on their own projects which might include writing a poem or song in their traditional languages, or beginning to compile a dictionary. At the end of the week, each person presents their project to the larger group. The Breath of Life workshop has provided descendants of North American Indian tribes with the tools to produce dictionaries out of the silence of archives, libraries, and extinct languages. It is being replicated else where in the world, e.g. early this year there was a Breath of Life workshop in Outback Australia and in the Canadian Arctic Archipelago of Nunavut.

6 Lexicography as a Means of Skills Transfer and Capacity Building

As seen with the Iquito Dictionary Project, the BOLD initiative, and the Teop Dictionary Project, the advent of language documentation as a field in itself has opened new opportunities for ensuring that dictionaries of endangered languages are community-focused and collaborative. New technologies and software allow dictionaries to imbed sound, video, and texts. They also allow multi-user access during the compilation process i.e. indigenous dictionary makers are jointly able to edit dictionaries with linguists living elsewhere in the world, thereby forming a dictionary team that can simultaneously work on the dictionary from different parts of the world. Such web-based collaboration is possible via an open-source software application called *Wesay*, produced by Summer Institute of Linguistics (SIL) in Papua New Guinea and Thailand. Intended for rugged low-power hardware, such as notebooks, *Wesay* specially caters

to the needs of indigenous dictionary makers by providing them with a simple and easy interface that requires minimal training. The software was developed especially for the speakers of endangered languages so that they can create their own dictionaries.

There are currently efforts to put this dictionary-making software on thousands of laptops being distributed to the world's poorest children via the One Laptop Per Child (OLPC) program. This initiative provides each child with a rugged, low-cost, low-power, connected laptop which is able to take photos and record sound. The software and learning materials currently provided on the laptops are in the dominant languages of the regions (e.g. mainly in Spanish or English) which of course increases the risk of endangering indigenous languages. Hence, there are currently initiatives to add Wesay dictionary-making software, along with lexicographic pedagogical materials, so that speakers of endangered languages in countries such as Peru, Rwanda, and Uruguay can document their languages and be introduced to dictionary-making activities at school level. It is hoped that classes will be able to compile their own mini dictionaries of community languages, thereby not only recording the languages but also increasing the children's use and pride in them.

The most sophisticated new dictionary-making technologies which enable speech communities to be involved in the documentation of their own languages are called Lexus and ViCoS. They are web-based tools by which lexicographers, both in the field and outside the field, can create (simultaneously) dictionaries that include sound, video, and immediate links to the relevant video segment where any word occurs. They also allow a dictionary to capture the indigenous view of the world by including a kind of visual thesaurus that presents indigenous semantic networks, i.e. networks that display the way speakers order and conceptualize semantic categories. These are particularly useful for communities that are not largely literate, and for dictionary users who rely more on visual and auditory than textual features. For example, in Gaby Cablitz's dictionary of the Polynesian language, Marquesan, a user can look up the meaning of a verb and see it in action. A user can look up the entry *kae,* a transitive verb meaning 'to cut or split off bark of a trunk or branch with a knife', and press a video to see how *kae* is performed.

The advent of documentary linguistics has encouraged lexicographers to integrate documentary materials into the text so as to create multimedia dictionaries which are more like cultural encyclopedias in their range. And, as we saw in the Yurok Dictionary, multimedia dictionaries can also

combine new lexical data with older archive material, allowing diachronic perspectives.

The inclusion of multimedia materials, and the desire for dictionaries of endangered languages to include socio-cultural information, opens the lexicographer to new considerations of ethical issues. The interests of the speakers are primary in the lexicographer's mind. In addition to negotiating extra issues with the speech community such as informed consent, payment for language consultants, and sharing outcomes, lexicographers of endangered languages must be mindful of cultural sensitivities surrounding the material they are documenting, i.e. access to sacred songs, taboo words, or the voice or image of Elders who may soon be dead (and whose name, voice, or image must not be uttered, heard, or seen for a certain period of time). Hence, many parts of the Yurok Dictionary are password protected. During dictionary work by Marina Chumakina on Archi, a north-east Caucasian language spoken by 1200 people in southern Dagestan, Russia, sound files were recorded for every word in the dictionary by member of the community. At the end of the project, it became apparent that in such a small community, where everyone knew each other's voice, the speaker was embarrassed that the rest of the community would hear her saying words considered taboo, such as intimate parts of the body. She asked for those files to be excluded, and her wish was respected. Similar issues surround illustrative sentences based on recorded speech that includes gossip or private stories which would be easily recognized within small speech communities.

Software such as Lexus and Wesay enable a dictionary to be compiled over the internet in a wiki-like fashion, and software such as ViCoS and Protégé enable the speech community to have a linguistic resource linked to the dictionary that represents their own intuitions and ontologies. For example, the dictionary of Yami, a language of Taiwan, includes links to ontologies which represent indigenous semantic connections between fish names, e.g. the Yami tripartite distinction between edible fish for young men, edible fish for women, and edible fish for old men.[22] The Yami Dictionary used Protégé software to show the semantic connections between the fish, but there is other software available, the most well-known being Kirrkirr. Kirrkirr pioneered work in semantic networks and was developed originally to work with the Warlpiri Dictionary, an Australian Aboriginal language, published by Mary Laughren and David Nash in 1983. Since then the software has been developed further by scholars at

---

22   See Rau et al (2009).

Sydney University and Stanford. By creating a semantic network view, the lexicographer presents the user with a network in which words in the dictionary that are semantically related are connected together by coloured lines – each colour represents a different relationship e.g. same meaning or alternate forms. By creating a semantic domain view, the lexicographer presents the user with nested nodes that represent semantic domains. Given the current limitations of remote places (lack of electricity, computers, and internet access), these online ontology tools are still a little way off being used to their full potential, but they are certainly indicative of the direction in which field lexicography is heading.

One issue to consider with dictionary software is that of archiving, which is neither reliable nor guaranteed especially as software is updated and changed. Therefore some field lexicographers avoid dictionary-making proprietary software because they are concerned about the longevity and archiving of their data. For example, the datafiles of the Yurok Dictionary and the Hupa Dictionary are XML documents and the interface is run via an XSL style sheet. This is wise when you consider that dictionary work on an endangered language may be the final record of the language, so it is imperative that it is stored in ways that are flexible, enduring, and easily accessible.

7   Conclusion: the Impact of Language Documentation on Lexicography

The emergence of the field of language documentation in the past decade has clearly had an impact on dictionary writing. And this paper has provided a glimpse of dictionary projects around the world that are creating methodologies that might be relevant or insightful to European lexicographers. 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.

On the macro level, language documentation has increased creation of, and access to, innovative dictionary technologies. It has also increased the opportunity for lexicographers to engage in capacity building, transfer of skills, and empowerment of community members to share the responsibility of dictionary making. On the micro level, the impact of language documentation on lexicography is perhaps even more tangibly

obvious. These dictionaries of endangered languages comprise a wider inventory from a variety of speech genres, with sophisticated multimedia materials, and new ways of preserving cultural memory and representing semantic and cultural ontologies. Content is linked to learning materials which facilitate language revitalization so that the dictionary becomes more than just a means of language preservation; it becomes the catalyst and focus for living language. These dictionaries challenge traditional types of dictionaries because they are everything in one. They combine aspects of the learner's dictionary, historical dictionary, encyclopaedic dictionary, talking dictionary, pictorial dictionary, video dictionary, and visual thesaurus. Consequently, the field lexicographer wears many hats. Her/his lexicographic methods and practices incorporate aspects of all genres of dictionary writing, and her/his mode of dictionary compilation is collaborative in nature. This paper has presented ways that lexicographers around the globe are able to preserve, maintain, and revitalize endangered languages. While Europe created and shaped the art of dictionary writing as we know it today, the rest of the world is taking it in new directions.

> **Bibliography**

Austen, P. K. (2006). 'Data and Language Documentation.' In J. Gippert, N. Himmelmann and U. Mosel (eds.) *Fundamentals of Language Documentation* Berlin: Mouton de Gruyter, 87-112.

Beier, Christine and Lev Michael (2006). 'The Iquito Language Documentation Project: Developing team-based methods for language documentation.' In *Linguistic Discovery* 4(1).

Beier, Chris (2009). 'Implementing Community Involvement in Collaborative Language Documentation: The Iquito Case.' *First International Conference on Language Documentation and Conservation.*

Bowern, Claire (2007). *Linguistic Fieldwork* Palgrave Macmillan

Crystal, D. (2002). *Language Death* Cambridge: Cambridge University Press.

Everett, D. (2006). *Linguistic Fieldwork: A Student Guide* draft 3 April unpubl. ms.

Fishman, Joshua (1991) *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages* Multilingual Matters.

Frawley, William, Hill, Kenneth, and Pamela Munro (2002). *Making Dictionaries: Preserving Indigenous Languages of the Americas* Berkeley: University of California Press.

Harrison, K. D. and Anderson, Gregory (2006). 'Na(t)ive Orthographies and Language Endangerment: Two Cases from Siberia.' In *Linguistic Discovery* 4 (1).

Himmelmann, N. P. (1998). 'Documentary and Descriptive Linguistics.' In *Linguistics* 36:165-191.

Himmelmann, N. P. (2002). 'Documentary and Descriptive Linguistics.' In O. Sakiyama and F. Endo (eds.). *Lectures on Endangered Languages* 5 Kyoto: Endangered Languages of the Pacific Rim 5:37-84.

Hinton, Leanne (1997). 'Survival of Endangered Languages: The California Master-Apprentice Program.' In *International Journal of the Sociology of Language* 123:177-191.

Hinton, Leanne (2001). 'The Master-Apprentice Language Learning Program.' In Hinton and Hale (eds.). *The Green Book of Language Revitalization in Practice* San Diego: Academic Press: 217-226.

Hinton, Leanne and Ken Hale (eds.). (2001). *The Green Book of Language Revitalization in Practice* San Diego: Academic Press.

Krauss, M. (1992). 'The World's Languages in Crisis.' In *Language* 68:4-10.

Michael, Lev (2009). 'Developing Infrastructure for team-based language documentation and description: the Module-and-Seminar Model.' In *First International Conference on Language Documentation and Conservation* 1 (1):1-6.

Newman, Paul (1999). 'We have seen the enemy and it is us: the endangered language as a hopeless cause.' In *Studies in the Linguistic Sciences* 28(2):11-20.

Newman, Paul (2003). 'The endangered language issue as a hopeless cause.' In M. Janse & S. Tol (eds.). *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches* Amsterdam: John Benjamins, 1-13.

D. Victoria Rau, Meng-Chien Yang, Hui-Huan Ann Chang, and Maa-Neu Dong (2009). 'Online Dictionary and Ontology Building for Austronesian Languages in Taiwan.' In *Language Documentation and Conservation* 3 (2):192-212.

Warner, Natasha, Butler, Lynnika, and Quirina Luna-Costillas (2006). 'Making a Dictionary for Community Revitalization; the case of Mutsun.' In *International Journal of Lexicography* 19 (3):257-285.

Warner, Natasha, Luna, Quirina, Butler, Lynnika, and Heather Van Volkinburg (2009). 'Revitalization in a scattered language community: Problems and methods from the perspective of Mutsun language revitalization.' In *International Journal of the Sociology of Language* 198:135-148.

Woodbury, T. (2003). 'Defining Documentary Linguistics.' In Peter K. Austin (ed.). *Language Documentation and Description* vol. 1 SOAS: University of London: 35-51.