

# LINGUISTICS, LEXICOGRAPHY, AND THE REVITALIZATION OF ENDANGERED LANGUAGES<sup>1</sup>

Sarah Ogilvie: *Lucy Cavendish College, University of Cambridge* (svo21@cam.ac.uk)

## Abstract

There is greater awareness now than ever before that languages are dying at an alarming rate. Hence, there is an urgency among linguists and indigenous communities to document, describe, archive, and revitalize endangered languages. Dictionaries play an important role in this process and, within the last decade, field linguists have developed innovative lexicographic methodologies, policies, and practices from which all lexicographers can learn. These dictionaries are written for a varied audience; they use technology in new ways, draw on oral as well as written sources, incorporate pedagogic materials, and involve indigenous community members throughout their compilation. Dictionaries of endangered languages are no longer merely static records that preserve language but are now being developed as dynamic, multi-functional tools for language maintenance and revitalization.

## I. Introduction

In 1926, at the age of twenty-four, the controversial American anthropologist Margaret Mead (1901–1978) was on her first field trip to Samoa. Mead (1926) 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’.

This is surely 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 have traditionally entailed recording the language, transcribing the language, and writing a grammar of the language. Writing a dictionary of the language was often just a stepping stone towards the grammar, rather than a goal in itself. Most dictionaries of endangered languages are therefore compiled by linguists who are not trained 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 went to live on Cape York Peninsula in northern Australia to write a dictionary and grammar of an endangered Aboriginal language spoken by two fluent speakers.

Twenty years ago, field lexicography was lagging behind commercial lexicography on all levels. In the world of language description, there was barely any overlap between field linguists and commercial lexicographers. 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.

What is the potential relationship between linguists and lexicographers? There have been changes in the field of descriptive and documentary linguistics in the past decade that suggest that linguists might have something to teach lexicographers. What do lexicographers and linguists have to teach each other from the area of endangered languages?

This special volume of *IJL* is an attempt to begin this dialogue. It brings together six papers by linguists who have lived and worked in field locations, and written dictionaries of endangered languages. Their lexicographic work is exceptional because of its use and development of new technologies; its methodology of collaboration and capacity building; and its techniques for language learning and revitalization.<sup>2</sup> Andrew Garrett from University of California, Berkeley, discusses his innovative work on the online dictionary of Yurok, a Native language of northern California, which links the lexicon to a corpus of historical and contemporary texts, and includes pedagogic materials for new language learners. John Hatton from the Summer Institute of Linguistics in Papua New Guinea describes the lexicographic software he developed, which not only enables indigenous communities to write their own dictionaries but also allows them to collaborate remotely, either online or offline, with linguists in separate locations. Mamari Stephens from Victoria University, Wellington, and Mary Boyce from University of Hawaii explain the sensitive lexicographic considerations they faced when incorporating traditional customary Maori legal terms into a modern Maori Legal Dictionary. The paper by Gaby Cablitz, University of Kiel, provides excellent insights from her work on a dictionary of the endangered Marquesan languages in French Polynesia. In her attempts to document traditional cultural knowledge within the dictionary, Cablitz explores ways to balance the tension experienced by many field linguists: how to produce a dictionary that best suits the dual audiences of scholar and speech community. The final paper by Nicholas Thieberger, University of Melbourne, highlights the importance for lexicographers of conforming to

established standards when compiling dictionaries of endangered languages, thereby producing lexical databases that are technologically flexible and able to be accessed, reused, edited, and linked to additional media in the future.

As a body of work, this volume provides a first step towards an engagement between linguists and lexicographers. The papers present a cross-section of the kind of work being done by field linguists who are writing dictionaries of endangered languages, and who are engaging seriously with lexicographic issues associated with technology, collaboration, and language documentation and revitalization. This article situates the work in this volume by surveying this field as it has developed.

## 2. Language shift and the role of dictionaries in language documentation and revitalization

Endangerment depends on the degree of language shift. Twenty years ago, Joshua Fishman (1991) designed eight steps toward reversing language shift. The steps progressed from the ultimate goal of step 1 - making the language the language of national government - to the 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. In contrast, speakers of endangered languages 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 the Berkeley linguist Leanne Hinton who shifted the focus from the national to the domestic, from the ultimate goal of government use to the realization that languages must first be spoken at home by children if they have a chance of being spoken anywhere (Hinton 1997, Hinton and Hale 2001). This change in scholarship has affected in fundamental ways the approach to linguistic description as well as the nature, focus, and quality of documentation and revitalization programmes. It has also changed the nature, focus, and quality of dictionaries of endangered languages in ways that all in this discipline can learn from, regardless of whether the aim is to promote a language to national or domestic level.

## 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 6600 languages become extinct, and most of these will disappear without being adequately recorded (Krauss 1992, Crystal 2002: 19). Current language distribution shows that languages are

extremely unequal in their demographic distribution. For example, a majority (3586) of the world's languages are spoken by a tiny proportion (0.2%) of the world's population, while a minority (83) of the world's languages are spoken by the majority (79.5%) of the world's population (Harrison 2007: 14). In his book *Language Death*, David Crystal (2002: 19) suggested that one language dies on average every two weeks. 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 do a culture and a speech community's unique way of seeing and ordering the world.

#### 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. As 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.

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, as evidenced by Frawley et al. (2002). These changes have been accompanied by new trends in documentary linguistics and anthropology as priority research areas that deal with the principles, theories, and practices of documenting languages and cultures that are at risk (Himmelman 2002, Woodbury 2003, Austin 2006). In 1998, in a landmark article in the journal *Linguistics*, Nikolaus Himmelmann formally distinguished between language documentation and language description. The aims of language documentation were to record the primary data of language study: to transcribe spoken and written texts, to annotate them with metadata, and to archive them for

posterity. Language description, by contrast, was concerned with the secondary data of language study such as the 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, as explained in the articles by Cablitz and Hatton in this volume, linguists 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, however, are not only useful and functional texts, but emblems and tools of prestige which many communities use to boost their sense of identity and their political profiles.

For the lexicographer, the field situation often presents a complex set of challenges that has an impact on lexicographic policies and practices. As the papers in this volume will testify, an undocumented language presents challenges relating to dictionary audience, format, and compilation. Questions relating to audience include: are you writing for scholars or the speech community, and if the latter are they fluent speakers, semi-speakers, new speakers, children or adults? Decisions to be made on format include: will it be a print dictionary, web-based, or electronic with imbedded pictures, sound, or video? Will the dictionary be linked to learning materials (as discussed by Andrew Garrett in this volume)? What software and format will ensure longevity of the dictionary project, and the ability to edit and to update over time (as considered by Nicholas Thieberger in this volume)? Issues relating to compilation that must be negotiated are: what orthography and writing system will you devise? Will you be combining new material with legacy and historical materials (as discussed by Mary Boyce and Mamari Stephens in this volume)? How might you list words in a dictionary for a polysynthetic language? How will the compilation involve the speech community and help capacity building (as addressed by Gaby Cablitz in this volume)? What software will you choose to accomplish this (an issue dealt with by John Hatton and Nicholas Thieberger in this volume)? Will it facilitate simultaneous editing, both online and offline, and in and out of the field? All of these issues - the audience, format, and mode of compilation - will depend on region; the 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-focussed 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.

### 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, Menriq, Batek, Lenoh, Maniq, and Semnam (Burenhult and Wegener 2009). 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 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 (<http://www.pfed.info/wksite>), to install dictionaries of endangered Australian Aboriginal languages on mobile phones. Loaded on to a mobile phone via software called Wunderkammer, a Java ME MIDlet, each dictionary entry has a spoken pronunciation and many entries have pictures. Currently, one Australian Aboriginal language, Wagiman, is on mobile phone and further projects are underway for Tura, a language of the Ivory Coast, and Whitesands, a language of Vanuatu.

It has long been a tradition in field lexicography to order a dictionary according to semantic fields of cultural relevance. Often a dictionary would be

ordered and published both alphabetically and semantically. In recent years there has been a trend in endangered-language lexicography to produce small dictionaries of separate semantic fields (Mosel 2004, 2011). 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 with 5000 speakers 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 (for example, 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 audience 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 have that speech community wait years for the completion of a comprehensive dictionary, has the joint benefit of demonstrating the lexicographer’s commitment to language maintenance and revitalization in the community, and showing an 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 (2006), advocates a team-based and community-participatory approach to dictionary writing that helps in fast collection of data.<sup>3</sup> The research team comprises two or three community linguists and four to seven visiting linguists (professors and graduate students) who visit the field at the same time. In all projects of this sort, 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 twenty-five speakers all of whom are over the age of sixty-five, 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’ (Beier 2009: 4).

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 created 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 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.

For critically endangered languages (those with no child speakers), it is not only necessary to record the language quickly, but it is also 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.

An example of a flagship dictionary that facilitates revitalization is the Yurok Dictionary created by Andrew Garrett and colleagues at UC Berkeley. As Garrett describes in this volume, the Yurok Dictionary is available free online and its entries are linked to language memory tests and language-learning exercises with audio files. The dictionary is similar in structure to the *Oxford English Dictionary* (*OED*) 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 by linguists and anthropologists at different times throughout the twentieth century. A quotation paragraph within each entry contains illustrative sentences that are linked to the larger texts and sound files in which they originally occurred, and users see a picture of the original speaker.

There is one important difference between the historical examples in the Yurok dictionary and those in the *OED*. The Yurok quotations 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. Inclusion policies in commercial dictionaries which are based on the number of citations from written sources get increasingly difficult to defend as technology improves our ability to capture, hear, verify, and reproduce natural speech in natural contexts. This may



be an area in which mainstream lexicography will follow innovations in field lexicography.

Lexicographers who write dictionaries for language revitalization 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 intensively together so that young members develop 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'.<sup>4</sup> The Africanist Paul Newman (1999, 2003), controversially criticizes this type of activity because he argues that lexicographers and academics should not become mere 'linguistic social workers' who waste their skills and time on the 'hopeless cause' of language revitalization. There are numerous examples of lexicographers around the world who demonstrate that this is not a hopeless cause, successfully negotiating a balance between dictionary work and revitalization work. For dictionaries written with revitalization as one of the outcomes, many would argue that dictionary work and revitalization work are inseparable.

One proven and successful methodology for bringing native speakers together with language learners is the Master-Apprentice Program, originally devised by Leanne Hinton (1997, 2001), Nancy Richardson, and Mary Bates Abbott for revitalization of Californian languages. 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 the language learners or lexicographers (the Apprentices). Hence, while learning the language from the Master, the lexicographer 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 (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: their attitudes towards the language in general and their willingness for inter-generational language transmission.

Activists for the preservation of endangered languages often stress the urgency of capturing and saving languages before they disappear, claiming it is literally a matter of life or death. Or is it? The logical extreme of dictionaries for revitalization, of course, are those that are written without direct contact with any speakers, from extinct languages. It is possible to revive a language from written sources alone (take, for example, 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 had no idea that their descendants would be using his wordlist to revive the language on Long Island in 2010.<sup>5</sup> Hence, as Nicholas Thieberger stresses in his paper in this volume, it is vital for lexicographers of endangered languages to create their lexical databases in forms that will endure and be readily accessible for future research and use.

An important issue for lexicographers of endangered languages to consider when choosing 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 software because they are concerned about the longevity and archiving of their data, perhaps choosing to create instead simple datafiles that are XML documents and an interface that 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 for future researchers and language learners.

Indeed, the current work of lexicographers of endangered languages will surely provide materials for language programs of the future. The exact shape, 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 users of their 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.'<sup>6</sup>

Similarly, the academic Natasha Warner and Mutsun tribal member Quirina Luna are compiling a dictionary of Mutsun, the language traditionally spoken south of San Francisco, California. It has been extinct, or 'dormant' as Warner et al. (2006: 259, 2009) prefer to describe it, since 1930, but they 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'. Luna has learnt the language and has taught it to her six children, the youngest of whom is four years old and only speaks Mutsun, no English. 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 collected 36,000 pages of notes on Mutsun 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 (Warner et al. 2006) were also faced with the task of inventing new Mutsun terms for the modern word, e.g. restaurant 'ammamsa' = eat + locative nominalizer.

The Mutsun dictionary initiative and the Unkechaug revitalization efforts came out of a workshop called *Breath of Life*, organized by Leanne Hinton, the same professor at UC Berkeley who developed the Master-Apprentice program. Every two years, the *Breath of Life* workshop brings sixty people who identify as Native American 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 linguists or lexicographers. They spend the week receiving intensive training each morning in the basics of linguistics and lexicography. Each afternoon, the participants are shown how to access the rich linguistic and anthropological archives housed at UC Berkeley, and each evening they work on their own projects which might include writing a poem or song in their traditional languages, or beginning to compile a dictionary.<sup>7</sup> At the end of the week, each person presents their project to the larger group. The *Breath of Life* workshop has provided descendants of Native American 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: recently there were *Breath of Life* workshops in Outback Australia, in the Canadian Arctic Archipelago of Nunavut, and at the Smithsonian Institute and Library of Congress in Washington.

## 6. Lexicography as a means of skills transfer and capacity building

As seen with the Iquito Dictionary Project, the Teop Dictionary Project, and the dictionary of the Marquesan languages described by Gaby Cablitz in this volume, the advent of language documentation as a field in itself has opened new opportunities for lexicographers to ensure that dictionaries of endangered languages are community-focussed and collaborative. New technologies and software allow dictionaries to imbed sound, video, and texts. They also allow multi-user access during the compilation process such that 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. John Hatton's article in this volume shows how such collaboration is possible via an open-source software application called Wesay which was developed by Hatton and colleagues at the Summer Institute of Linguistics (SIL) in Papua New Guinea. It is an indigenous equivalent of other SIL dictionary-making tools such as Shoebox, Toolbox, and Fieldworks Language Explorer (FLEX) which are most often used by field linguists. Intended for rugged low-power hardware, such as notebooks, Wesay especially caters to the needs of indigenous dictionary makers by providing them with a simple and easy interface that requires minimal training (Albright and Hatton 2008). The software was developed especially for the speakers of endangered languages so that they can create their own dictionaries and collaborate with dictionary team members who live in different parts of the world or have different degrees of reliable internet connectivity.

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. As the paper by Gaby Cablitz in this volume exemplifies, these are web-based tools that enable lexicographers, both in the field and outside the field, to create (simultaneously) dictionaries that not only include sound, video, and immediate links to the relevant video segment where any word occurs, but also allow a dictionary to capture the indigenous view of the world by including a kind of visual thesaurus that presents indigenous semantic networks, that capture the way speakers order and conceptualize semantic categories. Although they require technological expertise to compile, the final products can be 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 Cablitz's dictionary of the Marquesan languages in French Polynesia, a user can look up the meaning of a verb and see it in action (Cablitz et al. 2007 and this volume). At the entry for *kae* for example, which is a transitive verb meaning 'to cut or split off bark of a trunk or branch with a knife', the user can 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 legacy and 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 best 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 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, some parts of the Yurok Dictionary are password protected. During dictionary work by the linguist 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 a 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, 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 ethnobiologies and semantic connections between fish names such as the Yami tripartite distinction between edible fish for young men, edible fish for women, and edible fish for old men (Rau et al. 2009). 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 (Manning et al. 2001). 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 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 that field lexicography is heading.

## 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. 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. Their lexicographic methods and practices incorporate aspects of all genres of dictionary writing, and their 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.

## Notes

1 A version of this paper was given as a keynote address at the XIV Euralex International Congress, Leeuwarden, The Netherlands in 2010.



2 Most of these papers resulted from a Colloquium on Dictionaries and Endangered Languages organized by the author at the Second International Conference on Language Documentation and Conservation at the University of Hawaii in February 2011.

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

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

5 See 'Indian Tribes on Long Island Go in Search of Their Lost Languages' *New York Times* 6 April 2010 C5.

6 Robert Hoberman quoted in 'Indian Tribes on Long Island Go in Search of Their Lost Languages' NYT 6 April 2010 C5.

7 At the Breath of Life workshop at UC Berkeley in 2010, I collaborated with Quirina Luna to create an online phrasebook of Mutsun so that people who identify as Mutsun can learn the language from anywhere in the world. It can be accessed at <http://aicls.org/breathoflife/projects/mutsun/>.

## References

- Albright, E. and J. Hatton. 2008.** 'Wesay, a Tool for Engaging Communities in Language Documentation.' *Documenting and Revitalizing Austronesian Languages*, 1: 189–201
- Austin, 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. 2009.** 'Implementing Community Involvement in Collaborative Language Documentation: The Iquito Case.' *First International Conference on Language Documentation and Conservation*.
- Beier, Christine and Lev Michael. 2006.** 'The Iquito Language Documentation Project: Developing team-based methods for language documentation.' *Linguistic Discovery*, 4(1).
- Bowern, Claire. 2007.** *Linguistic Fieldwork*. London: Palgrave Macmillan.
- Burenhult, N. and C. Wegener. 2009.** 'Preliminary Notes on the Phonology, Orthography and Vocabulary of Semnam (Austroasiatic, Malay Peninsula).' *Journal of the Southeast Asian Linguistics Society*, 1: 283–312
- Cablitz, G., J. Ringersma and M. Kemps-Snijders. 2007.** 'Visualizing Endangered Indigenous Languages of French Polynesia with LEXUS.' *Proceedings of the 11th International Conference on Information Visualization*, 409–414
- 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, Kenneth Hill and Pamela Munro. 2002.** *Making Dictionaries: Preserving Indigenous Languages of the Americas*. Berkeley: University of California Press.
- Garrett, Andrew, Juliette Blevins and Lisa Conathan (compilers) 2005.** *Preliminary Yurok dictionary*. Berkeley: Yurok Language Project, Department of Linguistics, University of California.
- Harrison, K. D. 2007.** *When Languages Die*. Oxford: Oxford university Press.
- Himmelmann, N. P. 1998.** 'Documentary and Descriptive Linguistics' *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.' *International Journal of the Sociology of Language*, 123: 177–191.
- Hinton, Leanne. 2001.** 'The Master-Apprentice Language Learning Program.' In Hale Hinton (ed.), *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.' *Language*, 68: 4–10
- Manning, C. D., K. Jansz and N. Indurkha. 2001.** 'Kirrkirr: Software for Browsing and Visual Exploration of a Structured Warlpiri Dictionary.' *Literary and Linguistic Computing*, 162: 123–139.
- Mead, Margaret. 1926.** 'Letter from Margaret Mead to Franz Boas, 16 Jan, 1926.' *Correspondence Between Margaret Mead and Franz Boas Exchanged During Mead's 1925-26 Samoan Research Project (and related material)* <<http://sociology.uwo.ca/mead/>>.
- Michael, Lev. 2009.** 'Developing Infrastructure for team-based language documentation and description: the Module-and-Seminar Model.' *First International Conference on Language Documentation and Conservation*, 1.1: 1–6.
- Mosel, Ulrike. 2004.** 'Dictionary Making in Endangered Speech Communities.' In Peter K. Austin (ed.), *Language Documentation and Description*, 2: 39–54.
- Mosel, Ulrike. 2011.** 'Lexicography in Endangered Language Communities.' In P. K. Austin and J. Sallabank (eds), *Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 337–353.
- Newman, Paul. 1999.** 'We have seen the enemy and it is us: the endangered language as a hopeless cause.' *Studies in the Linguistic Sciences*, 28.2: 11–20.
- Newman, Paul. 2003.** 'The endangered language issue as a hopeless cause.' In M. Janse and S. Tol (eds), *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches*. Amsterdam: John Benjamins, 1–13.
- Rau, D. Victoria, Meng-Chien Yang, Hui-Huan Ann Chang and Maa-Neu Dong. 2009.** 'Online Dictionary and Ontology Building for Austronesian Languages in Taiwan.' *Language Documentation and Conservation*, 3.2: 192–212.
- Warner, Natasha, Lynnika Butler and Quirina Luna-Costillas. 2006.** 'Making a Dictionary for Community Revitalization; the case of Mutsun.' *International Journal of Lexicography*, 19.3: 257–285.
- Warner, Natasha, Quirina Luna, Lynnika Butler and Heather Van Volkinburg. 2009.** 'Revitalization in a scattered language community: Problems and methods from the perspective of Mutsun language revitalization.' *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.