elpublishing

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 2*. Editor: Peter K. Austin

Dictionary making in endangered speech communities

ULRIKE MOSEL

Cite this article: Ulrike Mosel (2004). Dictionary making in endangered speech communities. In Peter K. Austin (ed.) *Language Documentation and Description, vol 2*. London: SOAS. pp. 39-54

Link to this article: http://www.elpublishing.org/PID/019

This electronic version first published: July 2014

Dictionary making in endangered language communities

Ulrike Mosel

1. Introduction

This paper discusses a number of problems which are characteristic of lexicographic work in short-term language documentation projects and addresses the following issues: co-operation with the speech community, the selection of a dialect and the challenge to produce a useful piece of work meeting the usual academic standards of lexicography in spite of limited resources of time, money and staff, and the fact that the language under study is not well researched, the linguist does not have a thorough knowledge of the language and the local assistants do not speak the lingua franca fluently. Drawing on my experience with dictionary projects in Western Samoa and Papua New Guinea, I deal in particular with orthography, the compilation of wordlists and the writing of dictionary entries. A brief discussion of the organisation of workshops and a critique of the notion of capacity building concludes the paper.

2. Making dictionaries

Dictionaries of endangered languages are different from those of major languages in many aspects, and, of course, they themselves are not all of the same kind, but as diverse as other dictionaries are with regard to their size, coverage, and design. The main differences between ordinary commercial dictionary projects and those for endangered languages are that the latter are typically non-profit enterprises with limited resources of time, money and staff, and that the linguist who is responsible for the project is not a native speaker of the language.

Dictionaries of minority or endangered languages are often compiled by a single person, for instance a teacher or a missionary who lives in the community[1], or by linguists or anthropologists regularly visiting the speech community over many years, either as a part or a by-product of their research projects. Lexicography of this kind only receives acknowledgement from a few specialists, in most cases neither linguists nor the general public take any notice. This will hopefully change now. Thanks to the growing awareness of the endangerment of languages and cultures, language documentation projects have now been initiated by research institutions and funding agencies in increasing numbers[2], and lexicographic work will be part of language documentations though it will not necessarily result in dictionaries.

[1] Walker and Wilts (1987) give an interesting account of lexicographers of the North Frisian language spoken in Schleswig-Holstein, Germany.

[2] gbs bulletin, Mitteilungen der Gesellschaft für bedrohte Sprachen e.V., http://www.uni-koeln.de/gbs.

Ulrike Mosel (2004) Dictionary making in endangered language communities. In Peter K. Austin (ed.) *Language Documentation and Description, Vol 2*, 39-54. London: SOAS.

The present paper[3] focuses on dictionary projects, which are part of language documentation projects, or are carried out under similar conditions. Such dictionary projects on not well-researched languages are typically limited to rather short periods of time – short in comparison to projects on major languages, which may last decades.

The dictionary team usually consists of a linguist and a few native speakers from the endangered language community. Typically the linguist does not have a thorough knowledge of the language under investigation, the native speakers are not trained in linguistics; and to complicate things, both parties may not be fluent in the lingua franca they share as a means of communication.

Here, I want to discuss a number of problematic aspects of making dictionaries for endangered languages in short-term projects: setting the goal of the dictionary project, the time factor, the problem of giving preference to one dialect or variety of a language, the orthography, the question of how much grammatical information is necessary, the compilation of wordlists and the writing of dictionary entries. I suggest a variety of problem solving strategies, but as my personal experience as a lexicographer is limited to only two languages in the South Pacific, the Polynesian language Samoan, and Teop, an Austronesian language spoken in Bougainville, Papua New Guinea[4], these strategies may not work equally well in other parts of the world.

3. Setting the goal

When writing a dictionary for an endangered language, it is not enough just to say we want to document the lexicon of the language for future generations and generate a dictionary automatically from a text corpus. Rather, as with any other dictionary project, one has to identify the prospective users of the dictionary and what they will use the dictionary for. Being compiled in close co-operation with the speech community, the dictionary should serve the needs and interests of both the speech community and the academic community of linguists, anthropologists etc. Consequently, an electronic database, which seems to be the best medium for academic purposes, must be accompanied by a printed version for speech communities, without access to modern technology.

[3] I am thankful to Bruce Connell and Dafydd Gibbon for comments on a preliminary version of this paper.

[4] Samoan is not an endangered, unresearched language and the dictionary projects I was involved in were monolingual, but my experiences there helped me to develop strategies to deal with the time problem and to learn to work in a team with indigenous people. Teop, on the other hand, is a not well-researched endangered language which provides all sorts of orthographical, grammatical and semantic problems. The team working on the documentation consists of Ruth Saovana Spriggs, who speaks Teop as her first language, and Jessika Reinig and Marcia Schwartz and myself; cf. www.mpi.nl/DOBES. I am most grateful to Ruth Saovana Spriggs who introduced me into her language and her speech community and to Ruth Siimaa Rigamo my host and Teo teacher in Hiovabon, Bougainville.

Only recently have linguists become aware of ethical issues and questions such as: what does a fieldworker, or in our case a language documentation project, owe to the speech community as a proper acknowledgement of their contribution? What are their intellectual property rights (Newman and Ratliff 2001:9, Hinto and Weigel 2002)? From this perspective the individual local dictionary makers and the language community have the right to get copies of the dictionary in a form and with a content they appreciate. For most endangered language communities of the so-called third world, this means a printed version of the dictionary. This approach may also have some implications for the orthography, the selection of words (no taboo words), the macro-structure (strict alphabetical order or nesting), the micro-structure (not too much linguistic information in the entries) and the layout (large print). Conflicts between linguistic standards and user-friendliness as defined by the local dictionary makers[5] might be solved by producing two editions of the dictionary, one for academic purposes and one for the speech community.

In contrast to bilingual dictionaries of major languages, a dictionary of an endangered language does not serve as a tool for translation or foreign language acquisition, but as a resource for research and as a repository of the language for the speech community (see Sperlich 1997:1). For both purposes it is useful to complement the dictionary with a thesaurus covering the greatest possible variety of semantic fields such as kinship terms, animal and plant names, terms relating to the natural environment, the material culture and the social structure as well as all kinds of activities, states of being and properties. Since compiling thematically organized wordlists is an important part of the dictionary project, the work on the thesaurus is integrated in the work on the dictionary and need not be too time-consuming. A thesaurus can be very useful for the development of teaching materials and other language maintenance measures.

4. The selection of the language variety

The language to be represented by the dictionary may be spoken in more than one variety. Since the lack of time does not allow covering all of them, one dialect has to be given preference over the others. Quite often it is just the dialect of those people who invited the linguists to stay with them; in other cases the representatives of the language community might make the decision. If the linguist has the opportunity to select a dialect, they should consider the following criteria:

- which dialect is the most vital and is used in the greatest range of speech situations?
- are there children or young people who still use the dialect?
- which dialect is the most widespread one?
- where do the linguists find the most co-operative people?

[5] As Hausmann (1989:5, 14) remarks, there is always the danger of academic lexicographers loosing sight of the practical aspects and user friendliness of dictionaries because scientists do not think practical.

- where are the best native language experts located?
- where are the best living conditions?

Careful consideration is necessary. The mere fact that one dialect or speech variety is chosen for the compilation of a dictionary might influence the sociolinguistic status of that speech variety, which would certainly have some impact on the future development of the language.

Choosing the most vital dialect and giving it the prestige of being the language used in the dictionary or even becoming the standard language may be the ultimate death sentence for other dialects. On the other hand, the choice of a less vital dialect means that the dictionary and the language documentation might not cover the greatest possible range of speech situations.

5. The time factor

Since the project will be constrained by limited resources of money, staff and time, it must be organized in such a way that even after a very short period the dictionary makers can produce a useful piece of work. Instead of planning a comprehensive dictionary which would take many years to finish, one should consider being less ambitious and search for alternatives. There are, as far as I can see, two alternatives, which can be combined: Corpus Based Dictionaries and Thematic Dictionaries. Similar to the dictionaries of Classical Latin or Biblical Hebrew, Corpus Based Dictionaries only contain those words, which occur in a particular corpus of texts. The disadvantage of these dictionaries is that their content solely depends on the topics of the texts and the more or less accidental choice of words of the speaker, but if you have a large corpus of texts, they will certainly contain the most common words of the language. Thematic dictionaries, on the other hand, only cover the words of selected subject areas, such as kinship terminology or house building, but they may lack even the most common words if they do not belong to the selected themes. The advantage of Thematic Dictionaries, however, is that within a very short period of time you can produce a small but comprehensive dictionary which meets academic standards, and is interesting for people of the speech community as well as for academics of various fields.

The first dictionary I was asked to organize was a monolingual Samoan dictionary for the Ministry of Youth, Sports and Culture in Western Samoa in 1994, which was funded by the Australian South Pacific Cultural Fund with 10,000 Australian dollars. How could we, i.e. a staff member of the ministry and I as his consultant, produce a monolingual dictionary with these scarce resources? Necessity is the mother of invention: our first project was a little booklet on Samoan architecture and furniture (Mosel and Fulu 1997) and was later to be followed by similar mini-dictionaries on food, boat building, fishing and other culturally important practices (see section 8.4).

Any kind of dictionary project will benefit from training one or more locals to help with the work. If they are literate, they can learn to write their own wordlists and dictionary entries, if they are not, they can perhaps record lists of words and their translations or the explanations of their meaning in the source language (SL) by interviewing other people[6]. Literate locals can later continue with their own dictionary work (Fulu 1997).

6. Orthographic matters

Most endangered languages are not written languages or do not have a standardized orthography. If the native speakers who assist the linguist are literate in another language, the linguist should ask them to co-operate in developing a standardized orthography which can be easily written on a computer. The decision of which orthography is to be used should be arrived at in workshops with the local dictionary makers. As the standardization of the orthography is often a political matter, it can be difficult, but it must not hold up the compilation and production of the dictionary with never ending debates. While the linguists should always keep in mind that there is no such thing as the perfect orthography and not insist on their suggestions when the local dictionary makers take a different view, the latter should understand that not having a standardized orthography, that alternative spellings for numerous words may make the compilation of the dictionary impractical[7].

Sometimes, however, alternative spellings cannot be avoided. In Teop, for instance, vowel length is distinctive and long vowels are distinguished from their short counterparts by repeating the vowel letter, e.g. *na* a tense marker, and *naa* 'I'. Since the phonology of Teop has not been investigated in detail yet, we are often not sure how variation in vowel length is to be interpreted. In such cases we give the spelling variant just after the head word, whereas in example sentences we rely on the intuitions of the indigenous dictionary makers and often have the vowel spelled in different ways.

From the point of view of many linguists it might appear unreasonable or even irresponsible not to do a thorough phonological analysis before starting the work on the dictionary. However, it should not be forgotten that the work on the dictionary of an endangered language and culture is under severe time pressure as old people who can give us the most valuable information may die suddenly. With regard to the cultural aspects of our work, vowel length is a negligible problem.

[6] I have not practised this method myself, as the Samoans and Teops I work with are literate.

[7] For a recent discussion on the problems of variation and standardisation in endangered Amerindian languages cf. Rice and Saxon 2002.

7. Grammatical information

Unless it is accompanied by a grammar, the dictionary should at least contain as much information on the grammar in the front matter as is necessary to fully understand the abbreviations used in the dictionary entries such as, for instance, those used for the different parts of speech and their subclasses (see below the section on dictionary entries).

8. Writing wordlists

The first step in actually writing the dictionary is making lists of words which are to become head words or run-ons in dictionary entries. There are three methods:

- translating wordlists in the lingua franca into the source language as is suggested in most field manuals, at least for the basic vocabulary;
- extracting words from a text corpus;
- eliciting words by techniques which encourage the dictionary helpers to produce wordlists without translation.

8.1 The flaws of translating prefabricated wordlists

In many fieldwork manuals[8] you find wordlists in English (or some other European language) which are supposed to help you collect the basic vocabulary by translating the English words into the dictionary source language. For two reasons this method has to be used with caution: firstly, wordlists based on European languages will not be representative for the lexicon of the source language and miss all culturally specific concepts, some of which may also be basic. Secondly, the list may contain words which do not have a translation equivalent in the local language. Even items of the most "basic vocabulary" like 'eat', 'drink' and 'sit' (Swadesh 1972) may be missing (Goddard 2000).

More dangerous than this are, however, the psychological aspects of the translation method. The native speaker interviewees might feel very embarrassed when they are asked to translate a word they do not understand, or even worse, a word which they cannot translate because they have forgotten the native language equivalent.

Words elicited by translation always need to be counterchecked by translating them back into English later, or by explaining their meaning. The meaning of the SL word may be broader or narrower than its English counterpart, and the words in either language may be polysemous in different ways so that their meanings only partly overlap (see 9.3).

[8] Kibrik 1977: 103-123, Samarin 1967, Vaux and Cooper 1999:37-49. A remarkable exception is Beekman 1975.

8.2 Extracting wordlists from a text corpus

This method has the advantage of providing the words in a natural context which can also be used as examples in the dictionary. As the sense of the word in this context is often not its only sense, this method has to be complemented with asking native speakers for other examples of the use of the word. As Mithun (2001:38) observes, "a substantial proportion of the most interesting vocabulary emerges only in spontaneous speech, in what speakers themselves choose to say in different contexts." Consequently, the text material collected in the documentation project should cover a wide range of different speech situations. Wherever possible, the linguist should encourage the native speakers to make part of the recordings by themselves without the linguist present in order to avoid foreigner talk (Mosel 1984:13).

8.3 Active Elicitation

The method of extracting wordlists from a text corpus should be combined with a method I call Active Elicitation and which can be employed for all sorts of linguistic data, not just wordlists. Active Elicitation means that the language helper is asked to create their own set of SL data without translating words or sentences. After having explained what they are interested in, the linguists ask the language helpers to find words for narrowly defined subject areas such as the names of the plants they grow in the garden, types of houses, colours etc. As for activities, you might, for instance, ask what people do when preparing a meal ('get some water', 'wash the vegetables', 'make a fire' etc.), or you give the native speakers a basic word of a particular semantic field, eg. the speech act verb 'say', and ask them to search for similar words ('whisper', 'murmur', 'shout', 'ask', 'answer' etc.)

Active Elicitation encourages language helpers to participate actively in the dictionary work and understand what is being done so that they can identify themselves with the project and eventually, if they are interested, join the team of dictionary makers after some practice.

8.4 Alphabetical vs. thematic approach

Since the entries of the dictionary are alphabetically ordered, many people, speakers of Europeans and non-European languages alike, may think that the writing of a dictionary and hence the writing of wordlists starts with the letter A. In fact many dictionary projects employed this alphabetical approach and some of them were never finished, but stopped somewhere in the middle of the alphabet. A dictionary covering only the letters A to K (like a famous dictionary for Classical Arabic) is not a very useful book. In the case of endangered languages for which we may not have any dictionary yet, the alphabetical approach could be disastrous should the dictionary work come to an end before completion, for whatever reason. A dictionary which stops in the middle of the alphabet will lack many frequent words of the language and will not

cover any particular subject area of the culture in detail. Furthermore, it is unrealistic to plan to complete a comprehensive dictionary in less than ten years on this basis.

A more advisable method is the thematic approach, which was already mentioned above. This approach will be most efficient not only for the compilation of wordlists, but also for the writing of dictionary entries. After having selected a few subject areas and compiled the wordlists, the entries for each subject area can be written in turn[9]. When compared to the alphabetical method, the thematic approach has a number of advantages:

- the project can produce a useful piece of work with in a very short time. A small dictionary comprising just one subject area or sub-domain such as the names of winds or the types of houses and their parts is more useful than a 'dictionary' only comprising words starting with the letter A, as it could, for instance, be a resource for primary school teachers or ethnographers.
- producing a useful piece of work also raises the motivation of local lexicographers to help with the production of subsequent small dictionaries.

Whether you work over a longer period in the community or only come once a year for a short time, you can never be sure that you can always work with the same people. The thematic approach gives you the opportunity to finish the work on one subject area or sub-domain with one team, which will be easier and result in more consistent work than when you work on one subject with different people. Also the chances that important words are not omitted are better. Another advantage of the thematic approach is that indigenous dictionary makers can work on their special field of interest and interview experts on certain subject areas (e.g. fishing, architecture, healing, etc.), which certainly reinforces their motivation.

Since time and financial resources are limited, the project has to set priorities and decide which subject areas they want to work on first. Asking the following two questions may help the selection of the first subject areas:

- which subject areas do the native speakers think are most important for their culture?
- which subject areas are most suitable for the lexicographical training of indigenous dictionary makers?

The most suitable subject areas are those in which the SL meaning of the words is easy to explain by the native speakers and easy to understand by the non-native linguist, for

[9] This method was successfully employed for a monolingual Samoan students' dictionary which was written by a group of Samoan teachers under my supervision. In 150 pages it covers terms of about 20 subjects which were part of the curriculum of year 7 and 8, for example words related to the natural environment, certain plants and animals, shark and bonito fishing, house building, social science, mathematics and science (Mosel and So'o (eds.) 2000).

example subject areas of the material culture such as clothing or food preparation rather than theology or traditional law.

8.5 Purism

With regard to the selection of headwords, the local dictionary makers may be purists and puritans and wish to exclude borrowed or obscene words. As for borrowed words, I suggest trying to convince them that those which are adapted to the structure of the language belong to it and consequently should have their place in the dictionary. Otherwise the dictionary does not represent the living language as the people use it. Obscene and other taboo words are a more difficult issue. The speech community may agree to include them in a special academic edition of the dictionary.

9. Writing dictionary entries

When it comes to writing the dictionary entries, the problems of dealing with an under-researched language become most obvious. We already mentioned the orthographic issues, but other problems, as far as we experienced them in our Teop dictionary project, relate to word boundaries, classification of word classes and the productivity of derivational processes. As these are certainly not uncommon in other projects on not well-researched languages, we will describe them here with examples from the Teop dictionary project and present our solutions for discussion.

9.1 Word classes

In nearly all modern dictionaries the headword is followed by an abbreviation which indicates the so-called part of speech or lexical class of the headword, e.g. *n.* for noun, *v.* for verb, and *adj.* for adjective. Thus in an English-German dictionary you find: chicken *n.* 'Huhn'. But this practice is problematic in the case of the Teop dictionary.

In Teop words that correspond to nouns, verbs and adjectives in English are not inflected. It is not clear yet whether distributional criteria, collocational restrictions or derivational morphology are sufficient to set up lexical classes which could justifiably be called nouns, verbs and adjectives. Noun phrases, verb complexes and adjectival phrases, however, can be identified because they are marked by particular functional particles[10] such as articles and tense, aspect and mood markers[11].

Because of this uncertainty of classification, we decided to use the abbreviations *v., n.,* and *adj.* not in the traditional sense of lexical classes, but as labels for syntactic functions. Thus

[10] Mosel 1999a, 1999b.

[11] Cf. Broschart (1997) on Tongan who convincingly showed that Tongan does not distinguish between nouns and verbs, or Dixon (1977) on the lack of a distinctive class of adjectives in many languages of the world.

moon 1. *n.*, 'woman'. 2. *v.* '(be, become) a woman'.

has to be read: the word **moon** means 'woman' when it occurs as the head of a noun phrase, and '(be or become) a woman' when it occurs as the head of a verb phrase, where the semantic components 'be' or 'become' can be ascribed to the semantics of particular tense, aspect and mood particles.

Another problem is reduplication and derivation by affixes. Since we do not know to what extent these processes are productive, we list them all in the entries of the simplex as run-ons. Fortunately, the number of affixes is small (not more than five), otherwise nesting would result in very long entries and have a negative impact on the user-friendliness of the dictionary. Words derived by prefixes are also included as headwords without grammatical information or translation, but with a cross-reference to the root.

9.2 The definition or translation

The traditional division of dictionaries into monolingual and bilingual types does not need to be strictly observed in dictionaries of endangered languages, because they are not primarily used for translation. In fact, for many headwords a translation into a European language is not sufficient, because it won't capture the concept of the source language. In these cases a translation should be accompanied by a definition, which, if space does not matter, can be given in the source language first, and then translated. Such bilingual definitions would:

- preserve the interpretation of the meaning by the native speakers; which in case of misunderstandings on the part of the linguist may be most valuable in the future
- show the semantics of the headword and its relations to other words of the language as understood by native speakers
- make the dictionary a resource for further linguistic and anthropological research and for teaching materials

Good examples of the usefulness of bilingual definitions are animal and plant names. A dictionary, that only givse translations, would not show the taxonomy of animal and plant names in the indigenous language and culture, and would not explain the semantic relations between generic and specific terms. In Samoan, for instance, *atu* 'tuna', *malie* 'shark', *tafolā* 'whale', and *laumei* 'turtle' are grouped together as the same kind of animal, which is called *i'a*. A bilingual dictionary would only give the translation of 'turtle', but would not explain that it belongs to the class of *i'a*, which is mostly translated as 'fish', but is defined as "animal living in salt or sweet water, giving birth

to living off-spring or laying eggs"[12] in the monolingual dictionary (Mosel and So'o 2000:19).

Another area where translations often are not sufficient to render the meaning of a SL word is emotions. One example we came across when writing the Teop-English dictionary was the word *buruburusu,* which expresses the feeling someone has when they see or touch something strange or frightening (like a toad or a snake) and when this causes goose pimples and makes one's body hair stand up. There is no single English word which would express all these aspects of *buruburusu,* so an explanation is more useful.

As already mentioned, a severe handicap in lexicography for endangered languages is the time factor. One reason why writing a good dictionary takes so much time is that writing translations and definitions is so time consuming. Therefore, the lexicographers may be forced to be selective with regards to the number of headwords they translate or define, or the amount of information they give.

Since our Teop-English dictionary has to be completed in 2005 (we started to work regularly on the dictionary in 2000), we decided not to give the translations or the scientific names of plants and animals in the dictionary, but only indicate to which species they belong and give their characteristics. For example: "a hard wood tree growing near the coast whose timber is used for making canoes". We hope that sometimes in the near future an ethno-biologist can do field work and identify plant and animal names.

9.3 Illustrative sentences

In spite of the limited time, each sense of a headword should be illustrated by at least one example. Preferably the example sentences should not be made up, but come from recorded texts. They should illustrate the grammatical constructions of the word and contain frequently occurring collocations.

While gathering data for the dictionary, it is very helpful to have native speakers on the team who write down several structurally and contextually different examples sentences for each word. Such sentences can reveal semantic and morphosyntactic properties of the words that would otherwise have been overlooked. In our Teop dictionary, for instance, *babanihi* and *matavus* were both translated as 'door' and hence regarded as synonyms, until one of our local lexicographers provided the example

O	*babanihi*	*no*	*matavus*	*paa taketau.*
'The	"door"	of.the	"door"	is loose'

[12] *O le meaola e nofo i le sami ma le vai. O isi e tautu'ufua a o ni isi i'a e fanafanau.*

and we realised that two senses of the polysemous English word 'door', namely 'door-panel' and 'door-way' (Cruse 1986:65) are rendered in Teop by the two distinct lexemes *babanihi* 'door-panel' and *matavus* 'door-way'. Some training will be necessary for this – writing examples to order is a very specialised and skilled task.

9.4 Idioms and proverbs

As far as the limited time permits, lexicalised phrases and patterns of expression should be included in the dictionary, because the native speaker's linguistic competence does not only encompass the phonology, grammar and lexicon, but also the phraseology of the language[13]. One might also wish to include idioms and proverbs, because they reflect the culture of a speech community more than any other kind of linguistic unit or phenomenon, even if the explanation of their meaning and use can be difficult and time consuming.

9.5 Etymology

Although many linguists are interested in the history of languages, the etymology of words must be ignored. Since the documentation of an endangered language as a living language takes priority, the reconstruction of its history has to wait.

10. Capacity building: apprenticeship and workshops

The outcome of any kind of fieldwork relies heavily on co-operation between the academic outsider linguists and their local counterparts, which in the first place requires emotional intelligence and social competence on both sides (Goleman 1998). But it is the academic linguists who bear the full responsibility for a smooth effective workflow as it is they who introduced a new activity – dictionary making – into the community and who know how much, and what kind, of work is involved[14].

10.1 Mutual apprenticeship

In the beginning of any dictionary project of a previously unresearched language and culture, I recommend working with no more than three local people. Considering their personal interests and skills, they can be trained on an individual basis for specific lexicographic tasks like compiling wordlists, writing definitions and example sentences, and proofreading. These people are not, however, just apprentices of the craft of dictionary making, but are at the same time the linguists' mentors, teaching them the language and leading them to an understanding of their culture. Consequently, capacity building must be understood as a two-way process and not a patronising attitude of providing local people with

[13] Pawley 1993, 1996.

[14] Svensén 1993, chapter 21, gives an overview of the various components and stages of dictionary projects.

linguistic know-how without considering the fact that linguists acquire invaluable knowledge and experience through our work with them. What fascinates me as a researcher most in dictionary work is – to be honest – not lexicography as such, but the chance to learn through the language more and more about the conceptual world of a culture that is remarkably distinct from my own.

10.2 Workshops

In the South Pacific, workshops are frequently conducted by foreign aid agencies and non-governmental organisations in order to disseminate information, knowledge and new technology. Such activities might lead a speech community to wish for a workshop on dictionary making. Since the organisation of workshops is time-consuming and expensive, the purpose and possible outcomes of a workshop need careful consideration. From my experience, three kinds of workshops are useful:

- introductory workshops for community representatives, local language experts and teachers to inform them about the work processes involved in dictionary making, prevent wrong expectations and set realistic aims, justify the presence of the linguist in the speech community, and help to recruit local lexicographers;
- workshops that discuss the form and content of the dictionary or the orthography in order to facilitate the general acceptance of the dictionary;
- workshops for teachers on the use of the dictionary or other lexicographical materials resulting from the project.

It is not advisable to run workshops on compiling wordlists or dictionary entries with the aim to gather materials quickly and thus speed up the production of the dictionary. Certainly, you can get hundreds of pages in a two-day workshop with twenty people, but it can take months of frustrating work to sort out and revise these materials.

When you consider running a workshop, you need to form a small planning committee to become aware of the speech community's expectations, discuss objectives and feasibility issues, and calculate costs of transport, stationery, food and accommodation. The committee will also inform you on what kind of rituals and traditions of public discourse you may have to observe, and assist in designing a program.

11. Concluding remarks

Compiling a dictionary for an under-researched language means making compromises. The first edition of dictionary of an endangered language will not be a perfect dictionary. But as long as the dictionary makers are aware of problems and state explicitly in the front matter what kinds of problems they had to deal with and what kind of compromises or solutions they decided on, the dictionary can become a good resource for future research and language maintenance measures.

But it is not only the product – the dictionary – that serves language maintenance, the process of compiling it also does, if it proceeds in a spirit of co-operation and mutual respect for each other's contributions to the project. Not only do the compiling of wordlists, the discussions of meanings and the writing of entries raise awareness of the uniqueness and value of the language, but also lead to a general understanding that the linguist is making an effort to learn and analyse the language. Lexicography then can be a valuable two-way learning experience for all parties involved.

12. References

Beekman, John (1975). Eliciting Vocabulary, Meaning and Collocation. In Alan Healey, *Language Learner's Field Guide*, 361-386. Ukarumpa (Papua New Guinea): Summer Institute of Linguistics.

Broschart, Jürgen (1997). Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs. *Linguistic Typology* 1: 123-165.

Cruse, D.A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Dixon, Robert M.W. (1977). Where have all the adjectives gone? *Studies in Language* 1: 19-80.

Fulu, Mose (1997). *O le mafuaala o upu o taumafa. O le tusi faamatala upu o le gagana Samoa.* Apia (Western Samoa): Ministry of Youth, Sports and Culture.

Goddard, Cliff (2001). Lexico-Semantic Universals. *Linguistic Typology* 5: 1-65.

Goleman, Daniel (1998). *Working with emotional intelligence*. New York: Bantam Books.

Hausmann, F.J. (1989). Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. In F.J. Hausmann, O. Reichmann, H.E. Wiegand, L. Zgusta (eds.), *Wörterbücher, dictionaries, dictionnaies. Ein internationales Handbuch zur Lexikographie. Handbücher zur Sprach- und Kommunikationswissenschaft*, 5/1: 1-19. Berlin, New York: de Gruyter.

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

Kibrik, A.E. (1977). *The Methodology of Field Linguistics*. The Hague: Mouton.

Mithun, Marianne (2001). Who shapes the record: the speaker and the linguist. In Paul Newman and Martha Ratliff (eds.), *Linguistic Fieldwork*, 34-54. Cambridge: Cambridge University Press.

Mosel, La'i Ulrike and Mose Fulu (1997). *O le fale*. Apia (Western Samoa): Ministry of Youth, Sports and Culture.

Mosel, La'i Ulrike and Ainslie So'o (eds.) (2000). *Utugagana. O le tusi fa'amatala 'upu fa'apitoa mo tamaiti ā'oga o le Tausaga 7 and 8 o Samoa*. Apia: Department of Education.

Mosel, Ulrike (1984). *Tolai Syntax and its Historical Development*. Pacific Linguistics B-92. Canberra: The Australian National University.

Mosel, Ulrike and Ruth Spriggs (1999a). Gender in Teop (Bougainville, Papua New Guinea). In Barbara Unterbeck and Rissanen (eds.), *Gender in Grammar and Cognition*, 321-349. Berlin: Mouton de Gruyter.

Mosel, Ulrike and Ruth Saovana Spriggs (1999b). Negation in Teop. In Even Hovdhaugen and Ulrike Mosel (eds.), *Negation in Oceanic Languages*, 45-56. München: Lincom Europa.

Newman, Paul and Martha Ratliff (2001). Introduction. In Paul Newman and Martha Ratliff (eds.), *Linguistic Fieldwork*, 1-14. Cambridge: Cambridge University Press.

Pawley, Andrew K (1993). A language which defies description by ordinary means. In William A Foley, (ed.), *The Role of Theory in Language Description*, 87-129. Berlin: Mouton de Gruyter.

Pawley, Andrew K. Grammarian's lexicon, lexicographer's lexicon: Worlds apart. In Jan Svartvik (ed.), *Words KVHAA Konferenser 36*, 189-211. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.

Rice, Karen and Leslie Saxon (2002). Issues of standardisation and community in aboriginal lexicography. In William Frawley and Kenneth C. Hill (eds.), *Making dictionaries. Preserving indigenous languages of the Americas*, 125-154. Berkeley: University of California Press.

Samarin, William, J. (1967). *Field Linguistics. A Guide to Linguistic Fieldwork*. New York: Holt, Rinehart , and Winston.

Sperlich, Wolfgang B. (ed.) (1997). *Tohi Vagahau Niue. Niue Language Dictionary. Niuean-English with English-Niuean Finderlist.*

Svensén, Bo (1993). *Practical lexicography*. Oxford, New York: Oxford University Press.

Vaux, Bert and Justin Cooper (1999). *Introduction into Linguistic Field methods*. München: Lincom Europa.

Walker, A.G.H. and O. Wilts (1987). Who writes a North Frisian Dictionary for whom and why? *International Journal of the Sociology of Language* 64: 107-120.

Linguistic study by speakers: efforts of an institute

E. Annamalai

Languages have been studied and grammars of them written in the described language itself from the times of the ancient Indian grammarian Panini. Classical Sanskrit remained a spoken language before and after Panini described it. In the past, the grammarian may have learnt the language as their first or second language or as a dialect, but spoke it before coming to describe it. This was true of tribal languages in India also. Murmu, a teacher from the Santali community, analyzed the phonological system of his language and devised an orthography called Ol Chiki in the nineteenth century. This common method of describing a language using the language itself changed with the advent of modern linguistics. The modern grammarian may not speak the language at all, or may speak it in fragments, while recognizing that speaking it with a reasonably good command makes one a better grammarian. The researcher learns to speak the language while doing the analysis, not preceding it, unlike most missionary grammarians, who first learnt the language from its speakers before describing it.

Another change brought by modern linguistics is that the metalanguage of grammatical description is different from the language described. This came about because the grammarian could not communicate well enough in the language to be described, the technical vocabulary of the metalanguage was not rendered into the language described, and the grammar was addressed to an audience outside the language community. Another explanation for the fact that the described and describing languages are different is that grammatical description requires specialized training in theory and methods and in the language that goes with them; this is typically only available to those few who choose linguistics to be their profession. The emergence of syntactic studies initiated by the Generative Grammar paradigm in linguistics made grammaticality judgments part of the data, however the metalanguage (except for English and some other major languages) and the audience of grammatical description, as defined in modern linguistics, remains unchanged.

A change in the situation was felt to be warranted in sociolinguistic studies in order to reduce the effect of the observer's paradox, but this did not find widespread acceptance in practice in the profession. Language documentation, however, provides a serious opportunity for change because of the enormity of the work to be done and the recognition of the need for the linguistic work to be relevant and useful to the language community. Language speakers are co-opted in linguistic work as partners of professional linguists. The gain is for both linguistics and the community simultaneously. The gain is larger for linguistics in certain kinds of linguistic work such as dictionary making (see Mosel, this volume), correlational studies of language

E. Annamalai (2004) Linguistic study by speakers: efforts of an institute. In Peter K. Austin (ed.) *Language Documentation and Description, Vol 2*, 55-58. London: SOAS.