

Lexicography in endangered language communities

ULRIKE MOSEL

Kiel University, Germany

1. TRYING TO SQUARE THE CIRCLE

This chapter discusses a number of problems which are characteristic of lexicographic work in language documentation projects and addresses the following issues: the cooperation with the speech community, the selection of a dialect and the challenge of producing a useful piece of work meeting the scientific standards of lexicography in spite of limited resources of time, money and staff. Drawing on my experiences with dictionary projects in Samoa and Bougainville, Papua New Guinea, I outline the differences between lexical data bases as they typically result from language documentation projects on the one hand and the kind of dictionaries minority speech communities want for educational purposes on the other, and then show how such lexical databases can be exploited for the design of dictionaries that both satisfy the needs of the native speakers and the interests of linguists. Since it is impossible to create a comprehensive dictionary in a language documentation project, I applied the so-called thematic approach, in which lexicographers work on particular semantic domains such as trees, architecture or fishing that were selected by the speech community because of their cultural significance. The result of this approach is a series of small dictionaries, so-called mini-dictionaries, which not only contain linguistic but also encyclopaedic information.

The main difference between dictionary projects for major languages and those for endangered languages are that the latter are non-profit enterprises, and that the linguist who is responsible for the project is in most cases 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 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. However, 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, and lexicographic work has been acknowledged as an important component of language documentation. (Coward & Grimes 2000, Frawley, Hill & Munro 2002, Haviland 2006)

The staff working on the dictionary team typically consists of a linguist and a few indigenous people of the endangered speech community. While the linguist does not have a thorough knowledge of the language under investigation, the native speakers are not trained in linguistics; and to complicate matters, both parties may not be fluent in the lingua franca they share as a means of communication. This situation, however, has begun to change, as more speakers

of minority languages get access to university education, summerschools and specialized workshops (Woodbury 2003).

This chapter discusses the following issues: the planning of the lexical database and the dictionary (§2), the compilation of word lists (§3), the writing of entries (§4), the transformation of the lexical database into a dictionary (§5), and capacity building in the speech community (§6). I suggest a variety of problem solving strategies, but as my personal experience as a lexicographer is limited to only two Austronesian Oceanic languages in the South Pacific, the Polynesian language Samoan and the previously unresearched Western Melanesian language Teop¹, these strategies may not work equally well in other parts of the world.

2. PLANNING THE LEXICAL DATABASE AND THE DICTIONARY

The central component of a language documentation project is the corpus of recordings with transcriptions and translations which is usually accompanied by a lexical database that is assembled during the process of transcribing and translating. Toolbox, the most widely used software for making dictionaries of previously unresearched languages (Coward & Grimes 2000) allows one to structure database entries like dictionary entries and export them into various formats, including Rich Text Format (RTF). Although a print-out of the exported lexicon in RTF has the layout of an ordinary dictionary, it significantly differs from any kind of dictionary. But before we discuss these differences and the strategies of transforming the lexical database into dictionaries in the fifth section, we will deal with those lexicographic issues that are relevant for both the lexical database and the dictionary. While this section focuses on the planning of the dictionary project including setting the goal (§2.1), the time factor (§2.2), the selection of a language variety (§2.3), orthographical matters (§2.4), and the question of how much grammatical information is necessary or desirable (§2.5), the third and the fourth sections give an overview of how to compile word lists and to write entries.

2.1 Setting the goal

In contrast to bilingual dictionaries of major languages, a dictionary of an endangered language does not primarily serve as a tool for translation or foreign language acquisition, but as a resource for research and as a repository for language revitalisation and teaching in the speech community. But before creating a dictionary on the basis of the lexical database one must - as in any kind of dictionary project - identify the prospective users and the purpose of the

¹ 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 of indigenous people. Special thanks go to my Samoan counterparts Mose Fulu and Ainslie So'o. The Teop team comprises so many people that they cannot be enumerated here, I am most grateful to all of them, especially to Ruth Saovana Spriggs who introduced me to her language and her people and to Ruth Siimaa Rigamu my host, best friend and teacher in Hiiovabon, Bougainville. The Samoan projects were funded by the Australian South Pacific Cultures Fund and AusAid, and the Teop Language Documentation Project by the German Volkswagen Foundation, cf. www.mpi.nl/DOBES/Teop.

dictionary. Being compiled in close cooperation with the speech community, the dictionary should serve the needs and interests of both the speech community and the academic community of linguists and anthropologists. Consequently, an electronic database, which seems to be the best media for academic purposes, must be accompanied by a printed version for speech communities who do not have access to modern technology (cf. Schwartz et al 2007) and by a dictionary or a series of mini-dictionaries.

Only recently linguists have become aware of ethical issues surrounding language documentation: What does a fieldworker, or in our case the researcher of a language documentation project, owe to the speech community as a proper acknowledgement of their contribution? What are their intellectual property rights? (Newman & Ratliff 2001:9, Hinton & Weigel 2002, Dwyer 2006, Bower 2008:148-169). From this perspective the individual indigenous dictionary makers and the speech community have the right to get copies of the lexicographic work in a form and of a content they appreciate, and this means for most endangered speech communities a printed version of the dictionary. The acknowledgement of indigenous property rights implies that all the conditions made by the community must be satisfied before the dictionary can be printed and published.

Respecting the community's rights may also have implications on 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 scientific interests on the one hand and userfriendliness as defined by the indigenous dictionary makers on the other can be solved by producing two editions of the dictionary, one for scientific purposes and one for the speech community. But if there is not enough time for both, the community dictionary should have priority.

2.2 The selection of the variety of language

The language to be represented by the dictionary may be spoken in more than one variety. In general, 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 speech community might make the decision. If the linguist has the opportunity to select a dialect, she or he should consider the following criteria: Which dialect is the most viable one and is used in the greatest range speech situations? Are there children or young people who still use the dialect? Which dialect is the most widespread? Where do the linguists find the most cooperative people? Where are the best native language experts? And 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 can make it the standard which would certainly have some impact on the future development of the language.

Choosing the most viable dialect and giving it the prestige of being documented in a 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

viable dialect means that the dictionary and the language documentation would not cover the greatest possible range of speech situations.

2.3 The time factor: small is useful

Since the project is constrained by limited resources of money, staff and time, the project 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 decades to be finished, one should be 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 by the speakers or writers. As there are no cooking recipes in the Bible, the vocabulary of food preparation, which is essential in any community, is underrepresented in the dictionaries of Biblical Hebrew.

Thematic dictionaries, on the other hand, only cover the words of the selected semantic domains such as gardening or house building and may lack even the most common words. The advantage of Thematic Dictionaries, however, is that within a very short period of time you can produce a short, but comprehensive dictionary, which meets scientific standards and is interesting for people of the speech community as well as for academics of various fields.

The first dictionary project 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 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 & Fulu 1997) and was later to be followed by similar mini-dictionaries on food (Fulu 1997), and other culturally important practices (cf. section 5).

2.4 Orthographical matters: working without a standard

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, they can cooperate in developing a standardized orthography. Decisions on orthographic matters should be made in close consultation with indigenous dictionary makers, for example in workshops (see section 6). As the standardization of the orthography is often a political matter, it can be difficult, but it should not delay the production or distribution of the dictionary by never ending debates. (Aoki 2002:295-297, Hinton & Weigel 2002:156-160, Seifart 2006) 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 indigenous dictionary makers take a different view, the latter should understand that not

having a standardized orthography will make the compilation of the dictionary cumbersome. (Rice & Saxon 2002)

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 yet been investigated in detail, 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 headword, 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 die one after the other. With regard to the cultural aspects of our work, vowel length is a negligible problem (see also Hinton & Weigel 2002:167-168).

2.5 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, for example, those used for the different parts of speech and their subclasses. But a list of abbreviations, e.g. *v.* 'verb' and *v.t.* 'transitive verb' is not sufficient. Since 'verb' and 'transitive verb' can mean very different things in different languages and in different grammars of the same language, the grammatical features of the word classes and subclasses should be briefly explained.

3. WORDLISTS

3.1 Headwords and subheadwords

The headword is the first word of a lexical entry. It serves as the keyword for all the information given in the entry. Consequently, a headword like *horse* not only represents a single lexeme, e.g. *horse*, with its two senses of 'animal' and 'piece of sports equipment' but also multiword expressions such as *horse sense* or *straight from the horse's mouth*. A derived lexeme, e.g. *quickly*, can either be itself a headword (cf. LDOCE 2005; OALD 2000) or be subsumed as a subheadword (also called secondary headword or run-on) under the headword of the root, e.g. *quick* (cf. COBUILD 1987, Coward & Grimes 2000:77-87).

When, as in the case of the Samoan causative morpheme *fa'a-*, the derivational morphemes are prefixes, the root-oriented approach may impede the search for derived lexemes because the causative *fa'a-mate* 'kill' is found under headword *mate* 'die' (Milner 1966:138). Therefore the Samoan monolingual school dictionary has all causatives as headwords with the result that the entries of the letter F cover 20% of the dictionary (So'o & Mosel 2000). In contrast, the

Teops followed the root-oriented approach and accommodated the causatives, e.g. *vaa-mate* ‘kill’, as a subheadword in the entry of the root, e.g. *mate* ‘die’. (Schwartz et al. 2007) Perhaps they are more aware of derivational processes than the Samoans although their language is morphologically similar. Some Teops even write the causative prefix as a separate word. In other words, the question of whether the root-oriented or the lexeme-oriented approach is preferable for a particular kind of dictionary cannot be answered by the linguist on the basis of the morphological structure of the language, since the userfriendliness of the one or the other approach depends on how much the speakers intuitively or consciously know about their language .

In principle the headword can be any kind of lexical item that the lexicographers consider as a useful keyword for the user. Such keywords are not necessarily lexemes but can also be inflected wordforms whose stems are difficult to recognize like *bought*, suppletive wordforms like *went*, or even bound morphemes like the Samoan and Teop causative prefixes.

3.2 *Bad words: a note on purism*

With regard to the selection of headwords, the indigenous dictionary makers may be purists and wish to exclude borrowed or obscene words. As for borrowed words, I would try to convince them that those which are adapted to the structure of the language belong to the language and consequently should have their place in the dictionary. Otherwise the dictionary would not represent the living language as the people use it. Obscene and other taboo words are a more difficult issue. Perhaps the speech community would agree to include them in a special scientific edition of the dictionary or in a database with restricted access. (Hinton & Weigel 2002:166) What kind of words a community does not want to include in the dictionary is unpredictable. Thus the compilers of the Hopi dictionary had to exclude any information of ritual. (Hill 2002:303)

3.3 *Writing wordlists*

There are three methods of compiling word lists

- translating word lists in the lingua franca into the indigenous language as is suggested at least for the basic vocabulary in many field manuals (cf. 3.3.1);
- eliciting words by techniques which encourage the dictionary helpers to produce word lists without translation (cf. 3.3.2);
- extracting wordlists from the corpus (cf. 3.3.3);
- participant observation (cf. 3.3.4).

In the very first phase of a language documentation project wordlists are compiled for a preliminary study of phonetics and phonology and the design of a practical orthography. (Crowley 2007:95-97) Later on they are mainly extracted from the corpus, whereas elicitation is only used to check and supplement the data from the corpus.

3.3.1 The flaws of translating prefabricated word lists

In many fieldwork manuals (cf. Abbi 2001:244-247, Bowerman 2008: 223-224, Kibrik 1977: 103-123, Samarin 1967:220, Vaux, Cooper & Tucker 2007:89-96) you find word lists in English which are intended to assist collecting the basic vocabulary by translating the English words into the indigenous language. For two reasons this method has to be used with caution: Wordlists based on a European or any other contact language – even a closely related one - will not be representative for the lexicon of the indigenous language and consequently miss many culturally specific concepts. On the other hand, the list may contain words, which do not have a translation equivalent in the indigenous language. Even items of the most “basic vocabulary” like ‘eat’, ‘drink’ and ‘sit’ may be missing (Goddard 2001). If you ask indigenous research assistants to translate an expression X of the contact language into their native language, then “X comes from a different linguistic system than the mother tongue of the person being asked.” (Grimes 2002:71) The meaning of the indigenous word may be broader or narrower than its counterpart in the contact language, and the words in either language may be polysemous in different ways so that their meanings only partly overlap.

Pawley, who in the sixties started compiling a dictionary of Wayan, a Fijian language, and used a dictionary of standard Fijian for his first elicitation, reports (2009:18): “Eliciting from lists can certainly yield quick results but my experience is that unless it is combined with a good practical knowledge of the language and careful checking it will leave a lot of errors.” Words, which have been elicited by translation, always need to be counterchecked by back-translation into the contact language, and by giving example sentences and explanations of their meaning. With respect to comprehensiveness, a useful tool are the extensive questionnaires on traditional technology, ethnobiology and anatomy in Bouquiaux and Thomas (1992:401-687), which also contain numerous illustrations.

Even more dangerous than the errors in semantic analysis are the psychological aspects of the translation method. The indigenous 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 indigenous equivalent.

3.3.2 Active Eliciting

The problems of the translation method can be avoided by a method I would like to call Active Eliciting. Active Eliciting means that the indigenous assistants are asked to create their own set of data without translating words or sentences. After having discussed the aims of the session, e.g. the compilation of a list of words that are suitable for combining them in phrases and clauses, and that in addition will be used for the study of sounds, the assistants choose a narrowly defined semantic domain such as food and list, for example, the names of vegetables, colour terms and words that have something to do with the preparation of meals (‘get some water’, ‘peel’, ‘cut’, ‘wash’, ‘make a fire’, ‘pot’ etc.). In this way each word of the list is associated with a particular context of the speech community’s

culture and thus naturally renders a concept of their language. Another way of eliciting lexemes is to give a basic word of a particular semantic domain, for example the speech act verb ‘say’, and ask for similar words, e.g. ‘whisper’, ‘murmur’, ‘shout’, ‘ask’, ‘answer’, etc. For further information on elicitation methods and a critical discussion see Haviland (2006:148-159) and Grimes (2002).

3.3.3 *Extracting wordlists from a text corpus*

The corpus based compilation of wordlists has the advantage that it provides the words in natural contexts. But as the sense of a word in a particular context is often not its only sense, this method has to be supplemented by asking native speakers for further examples that might reveal different senses. Furthermore, a text corpus compiled in a documentation project of three to five years is usually too small to cover a substantial part of the lexicon of the language so that the dictionary makers must resort to active eliciting.

3.3.4 *Participant observation*

Another method of obtaining natural data is listening to the people and noting down utterances that we encounter “in circumstances that we do not control.” (Grimes 2002:76) 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.”

4 WRITING ENTRIES

4.1 *The structure of the entry*

The lexical database provides for each entry and subentry a number of fields that in the process of the corpus analysis can be filled with grammatical, semantic and pragmatic information as well as illustrative sentences. For a detailed description of entries in databases and dictionaries of European languages see Atkins & Rundell (2008:100-101, 317-379) and Svensén (1993:210-218), for the building of a Toolbox database see Coward & Grimes (2000). The lexical entry starts with a field for the headword followed by fields for information on the homograph number, variants, pronunciation, part of speech and inflection. In the case of polysemous headwords like English *mouse* the lexical database provides a field for each sense number and subsequently, for each sense number separately, fields for grammatical information like 1. Plural *mice*, 2. Plural *mouses*, the meaning of each sense, e.g. ‘animal ...’ and ‘part of computer equipment ...’, citations from the corpus and examples created by native speakers with their references, and whatever additional fields the lexicographer considers as important for each sense.

The sequence of the senses of the headword can follow various principles. (Svensón 193-214) Ideally the most general meaning comes first and those that are restricted to certain contexts and are derived by metaphor or metonymy later. Thus in the entry of *atovo* in the Teop lexical database the sense ‘sagopalm’ precedes the sense ‘thatch made from sagopalm leaves.’ But note that sense

discrimination and the distinction between polysemy and homonymy is very difficult from both a practical and a theoretical perspective. For further information on this topic see Atkins & Rundell (2008:263-316), Cruse (1986:49-83), and Cruse (2000:104-142).

If a headword belongs to more than one part of speech category like *strike*, which functions as a verb and a noun, the part of speech field should be clearly marked off “to inform the user immediately that the entry contains more than one part of speech.” (Svensén 1993:210); cf. *strike* in the OALD 2000 and COBUILD 1987. Another possibility is to give each part of speech the status of a headword, cf. LDOCE 2003.

4.2 *Inflectional vs. derivational morphology*

When building a lexical database of a previously unresearched language, it may be impossible to decide whether a morphologically complex word, for example a word with reduplicated segments, is an inflected wordform or a derived lexeme because we do not know to what extent the respective morphological processes are productive and in certain contexts obligatory. In the Teop lexical database we therefore enter them as subheadwords. But before the database is transformed into a dictionary, the grammatical analysis of the language should have solved this problem in order to avoid unnecessarily complex entries.

4.3 *The definition*

The traditional division of dictionaries into monolingual and bilingual dictionaries 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 equivalent is not sufficient, because it would not capture the concept of the indigenous language. (Haviland 2006) In such cases the translation can be accompanied by a definition, which ideally is given in the indigenous language first and then translated. Such bilingual definitions would

- preserve the interpretation of the meaning by the native speakers, which reduces the danger of misunderstandings on the part of the linguist;
- show the semantics of the headword and its relations to other words of the language;
- make the dictionary a resource for further linguistic and anthropological research;
- can later be used for the development of a monolingual dictionary and teaching materials.

For each semantic domain you need to prepare a style guide that suggests what kind of information the definition should contain. For fishes this can be, for example, the habitat, size, colour, age and whatever the indigenous fishermen regard as important. (Coward & Grimes 2000: 137-152) In contrast to translations such plant and animal definitions could also show folk taxonomies.

In Samoan, for instance, *atu* ‘tuna’, *malie* ‘shark’, *mumua* ‘dolphin’, and *laumei* ‘turtle’ belong to the same kind of animals, which is called *i’a*. A bilingual

dictionary would only give the translation of ‘turtle’, but not explain that it belongs to the class of *i’a*, which is mostly translated as ‘fish’, but defined as ‘animal living in salt or sweet water, giving birth to living off-spring or laying eggs’² in the Samoan monolingual dictionary (So’o & Mosel 2000:19).

As already mentioned, lexicography for endangered languages is often severely restricted by considerations of time and money so that lexicographers are forced to be selective with regards to the number of headwords they translate or define and the amount of information they give. In the Teop lexical database we entered all lexemes we came across, but for various reasons we could not employ a biologist who would have identified the scientific names of animals and plants. Very often we could not even translate these names into English so that we had to resort to a definition that would provide the name of the superordinate class and typical characteristics, e.g. ‘a hard wood tree growing near the coast whose timber is used for carving canoes.’ Wherever possible, this kind of definition is illustrated by drawings and photographs.

4.4 Examples

In the database the example phrases and sentences are citations from the corpus that give evidence of the usage of the lexical units, i.e. the senses of the lexeme or lexicalized multiword expressions in their particular senses, see Cruse (1986: 23ff), Atkins & Rundell (2008:162f). Only for elicited lexical units will the native speakers create examples. Both types of examples serve the grammatical and semantic analysis of lexical units. In contrast, the examples of the dictionary reflect the result of this analysis and help the reader to better understand the salient grammatical and semantic features of the lexical unit. Since space does not matter in the lexical database, the lexicographers will collect everything that seems useful for the analysis, but when they transform the database into a dictionary, they must be selective and choose only one or two examples for each type of usage. Furthermore, many citations will prove unsuitable for dictionary examples because they are too complex, only comprehensible in their wider context and not representative like for example, the following citation from a legend: ‘The old woman hid the moon in her saucepan.’ Thus for practical reasons many citations need to be abridged, adapted to the format of example sentences or replaced by created representative examples.

Even if all lexical units of a headword seem to be well illustrated by citations from the corpus, it is advisable to ask native speakers to additionally supply examples. The created examples may illustrate more typical or different usages of the lexical unit and reveal misinterpretations. In the Teop lexical database, for instance, *babanihi* and *matavus* were both translated by ‘door’ and hence regarded as synonyms, until one of our local lexicographers provided the example

- (1) *O babanihi no matavus paa taketau.*
‘The door of.the door is loose.’

² *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 realized that the 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'. A good summary of functions of example sentences is found in Bartholomew & Schoenhals (1983:59-69).

4.5 Idioms and proverbs

As far as the limited time permits, lexicalized phrases and patterns of expression should be included in the dictionary, because the native speaker's linguistic competence not only encompasses the phonology, grammar and lexicon, but also the phraseology of a language. (Pawley 1992, 1993) 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, but the explanation of their meaning and use can be difficult. A classification of multiword expressions and a description of how they are treated in various types of English dictionaries is given in Atkins & Rundell (2008: 166-176, 222-225).

4.6 Etymology

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

5 TURNING THE LEXICAL DATABASE INTO MINI-DICTIONARIES

Since the entries of the dictionary are alphabetically ordered, many people think that the writing of a dictionary starts with the letter A. In fact many dictionary projects used 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 is not a very useful book. In the case of endangered languages, for which no previous dictionary exists, the alphabetical approach is disastrous if for whatever reasons the dictionary work comes to an end.

5.1 Advantages of the thematic approach

As already mentioned in section 2.3, the alternative to the alphabetic approach is the thematic approach, because a mini-dictionary can be produced in a rather short time and serve as a resource for teachers, linguists and ethnographers (cf. Mosel & Fulu 1997). Furthermore, the completion of such a booklet raises the motivation of indigenous lexicographers to continue the dictionary work by themselves once the professional consultant has left. (cf. Fulu 1997)

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 domain or subdomain with one team, which will result in a more

consistent piece of work than when you work on one domain with different people.

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. Furthermore, this approach bears an important advantage for the training of indigenous lexicographers: as a rule, specialized vocabulary is less frequently used than the core vocabulary, less polysemous and consequently easier to describe. (cf. Atkins & Rundell 2008:263) Because of its low frequency it is also the most endangered vocabulary.

The production of a non-commercial dictionary is expensive. It may not be possible to provide every teacher with a dictionary, but if every school receives a set of mini-dictionaries, the teachers can share this set. While one is using the mini-dictionary of architecture, the others may prepare their lessons with the fishing, gardening, tree or shellfish mini-dictionary. The animal and plant mini-dictionaries will also foster the awareness of biodiversity and may eventually contribute to the protection of the environment.

5.2 *The choice of themes*

Since time and financial resources are limited, the project has to set priorities. Two criteria seem to be a useful guide for the selection of the first semantic domains: Which domains do the elders and teachers consider as the most important ones for the transmission of their cultural knowledge to future generations? Which domains do you as a linguist or anthropologist regard as the most endangered ones?

Food preparation may be an excellent domain to start with because direct observation, photos and videos help the linguist to understand the terms denoting ingredients and activities, whereas the concepts of traditional law and cosmology would presuppose a deep understanding of the history and culture of the people.

A drawback of the thematic approach, however, is that some lexemes are polysemous like Teop *atovo* 1. 'sagopalm, 2. 'thatch of sagopalm leaves' so that their various senses would belong to different mini-dictionaries, e.g. the tree dictionary and the house dictionary. For the Teop mini-dictionaries we solved this problem by always giving the basic sense first, even if it belongs to a different semantic domain. Thus in the house dictionary the entry of *atovo* starts with the sense 'sagopalm'.

5.3 *Turning the lexical database into a mini-dictionary*

After having selected the semantic domains, the lexical units of the lexical database (i.e. the senses of a headword) are accordingly classified by entering the respective keywords into a special semantic domain field. If a lexical unit relates to more than one domain, the field can be filled with more than one keyword (for Toolbox see Coward & Grimes 2000: 26, 191). The semantic domain field can be hierarchally structured, e.g. *atovo* 'sagopalm': plant, tree; *taruvana* 'giant pandanus': plant, pandanus. (cf. Atkins & Rundell 2008: 182-184)

Then the lexical database is filtered so that for each semantic domain, e.g. 'plant', or subdomain, e.g. 'tree', a separate mini-database can be exported. These specialized mini-databases are then again exported and printed in a format that looks like the planned mini-dictionaries so that they can be discussed with experts and teachers of the speech community and accordingly be revised and supplemented. Eventually the project will produce a paper and an electronic version of the mini-dictionaries.

6. CAPACITY BUILDING: APPRENTICESHIP AND WORKSHOPS

The outcome of any kind of fieldwork heavily relies on the cooperation between the academic outsiders and their indigenous counterparts, which in the first place requires emotional intelligence and social competence on both sides. But it is the academic linguists who bear the full responsibility for a smooth effective workflow as it is they who introduce a new activity - dictionary making - into the community and who know how much and what kind of work is involved, see Svensén's overview of the various stages of dictionary projects from the planning phase till the final proofreading. (1993:236-249)

6.1 Mutual apprenticeship

In the beginning of the project I would recommend working with no more than three indigenous people. Considering their personal interests and skills, they can be trained on specific lexicographic tasks like compiling wordlists, writing definitions and example sentences, and proof reading on an individual basis. These people are not, however, just apprentices of the craft of dictionary making, but are at the same time the linguists' mentor teaching them the language and leading them to an understanding of their culture. Consequently, the notion of capacity building has to be rejected as reflecting a patronising attitude if we understand it exclusively in the sense of providing indigenous people with linguistic know-how without considering the fact that we acquire invaluable knowledge and experience through our work with them.

6.2 Workshops

In developing countries workshops are frequently conducted by foreign aid agencies and non-governmental organisations in order to disseminate information, knowledge and new technology. 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 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.

From our experience with the Samoan monolingual school dictionary it does not seem advisable to run workshops on compiling wordlists or dictionary entries. Certainly, you can get hundreds of pages in a two-days workshop with twenty people, but it will take months of frustrating work to sort out and revise these materials.

When you consider conducting 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 the costs of transport, stationary, food and accommodation. The committee will also inform you on what kind of rituals and traditions of public discourse you have to observe and assist in designing a program.

7. CONCLUDING REMARKS

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

But it is not only the product, the dictionary, that serves language maintenance. The whole process of making a dictionary, if done in the spirit of cooperation and mutual respect of each other's capacities, raises the awareness of the uniqueness and value of the language.

8. REFERENCES

8.1. Books

- Abbi, Anvita. 2001. *A manual of linguistic fieldwork and structures of Indian languages*. München: Lincom Europa.
- Atkins, B.T. Sue & Michael Rundell. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- Bartholomew, Doris A. & Schoenhals, Louise. 1983. *Bilingual dictionaries for indigenous languages*. Hidalgo, Mexico: SIL.
- Bouquiaux, Luc & Jacqueline Thomas (eds.). 1992. *Studying and describing unwritten languages*. Dallas: SIL.
- Bowern, Claire. 2008. *Linguistic fieldwork*. A practical guide. New York: Palgrave Macmillan.
- COBUILD *Collins Cobuild English Language Dictionary*. 1987. London: Harper Collins Publishers.
- Crowley, Terry. 2007. *Field linguistics*. A beginner's guide. Edited and prepared for publication by Nick Thieberger. Oxford: Oxford University Press.
- Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Cruse, D. Alan. 2000. *Meaning in language*. An introduction to semantics and pragmatics. Oxford: Oxford University Press.
- Frawley, William, Kenneth C. Hill & Pamela Munro (eds.). 2002. *Making dictionaries. Preserving indigenous languages of the Americas*. Berkeley etc.: University of California Press.
- Kibrik, A.E. 1977. *The Methodology of Field Linguistics*. The Hague, Paris: Mouton.
- LDOCE *Longman dictionary of contemporary English*. 2005. 4th edn. Harlow: Pearson Education limited.
- Milner, George B. 1966. *Samoan dictionary*. Oxford: Oxford University Press.
- OALD *Oxford Advanced Learner's Dictionary* 2000. Oxford: Oxford University Press.
- Samarin, William, J. 1967. *Field Linguistics. A Guide to Linguistic Fieldwork*. New York, etc: Holt, Rinehart, and Winston.
- So'o, Ainslie & La'i Ulrike Mosel 2000. *Utugagana. O le tusi fa'amatala 'upu fa'apitoa mo tamaiti a'oga o le Tausaga 7 & 8 o Samoa*. Apia, Samoa: Department of Education.
- Svensén, Bo. 1993. *Practical lexicography*. Oxford, New York: Oxford University Press.
- Vaux, Bert, Justin Cooper & Emily Tucker. 2007. *Introduction into Linguistic Field methods*. München, Newcastle: Lincom Europa. Eugene, Oregon: Wipf & Stock Publishers.

8.2. Articles in edited volumes, conference proceedings and working papers

- Aoki, Haruo. 2002. Writing a Nez Perce Dictionary. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries. Preserving indigenous*

- languages of the Americas*, 285-298, Berkeley etc.: University of California Press.
- Dwyer, Arienne M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 31-66. Berlin, New York: Mouton de Gruyter.
- Grimes, Joseph E. 2002. Lexical functions as a heuristic for Huichol. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries. Preserving indigenous languages of the Americas*, 70-85, Berkeley etc.: University of California Press.
- Haviland, John B. 2006. Documenting lexical knowledge. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 129-162. Berlin, New York: Mouton de Gruyter.
- Hill, Kenneth C. 2002. On publishing the *Hopi* Dictionary. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries. Preserving indigenous languages of the Americas*, 299-311, Berkeley etc.: University of California Press.
- Hinton, Leanne & William Weigel. 2002. A dictionary for whom? Tensions between academic and nonacademic functions of bilingual dictionaries. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries. Preserving indigenous languages of the Americas*, 155-170. Berkeley etc.: University of California Press.
- Mithun, Marianne. 2001. Who shapes the record: the speaker and the linguist. In Paul Newman & Martha Ratliff (eds), *Linguistic Fieldwork*, 34-54. Cambridge: Cambridge University Press.
- Mosel, Ulrike. 2006. Fieldwork and community language work. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 67-85. Berlin, New York: Mouton de Gruyter.
- Newman, Paul & Ratliff, Martha. 2001. Introduction. In Paul Newman & 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. 1992. 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 & Leslie Saxon. 2002. Issues of standardisation and community in aboriginal lexicography. In William Frawley, Kenneth C. Hill & Pamela Munro (eds.), *Making dictionaries. Preserving indigenous languages of the Americas*, 125-154. Berkeley etc.: University of California Press.
- Seifart, Frank. 2006. Orthography development. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel. *Essentials of language documentation*, 275-299. Berlin, New York: Mouton de Gruyter.

Woodbury, Tony. 2003. Defining documentary linguistics. In Peter K. Austin (ed.) *Language documentation and description*. Vol. 1, 35-51. London: Hans Rausing Endangered Languages Project.

8.3 Articles in journals

Goddard, Cliff. 2001. Lexico-Semantic Universals. In *Linguistic Typology* 5, 1-65.

8.4 On-line papers, reviews, dissertations and other kinds of publication

Coward, David & Charles E. Grimes. 2000. *Making Dictionaries*. A guide to lexicography and the Multi-Dictionary Formatter. SIL International Waxhaw, North Carolina.

http://www.sil.org/computing/shoebox/MDF_2000.pdf

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

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

Pawley, Andrew K. 2009. On the treatment of animal and plant names in bilingual dictionaries: lessons from Oceania, 28 pp., 11th International Conference on Austronesian Linguistics, Aussois, June 22-25 2009.

Schwartz, Marcia L., Ruth Saovana Spriggs, Ulrike Mosel, Ruth Siimaa Rigamu, Jeremiah Vaabero, & Naphtaly Maion 2007. *The Teop lexical database*. <http://www.mpi.nl/dobes/projects/teop>