

## Creating educational materials in language documentation projects – creating innovative resources for linguistic research

Ulrike Mosel  
*Kiel University*

In its first two sections this paper briefly discusses two models of language documentation projects: the hierarchical model, in which the language documentation corpus (LDC) serves as a resource for the development of educational materials (EMs), and the integrative model, which integrates the production of EMs into the LDC and makes them a resource for linguistic research. The third and the fourth section describe how the integrative model was applied in the Teop Language Documentation Project and what kind of linguistic research topics it provides.

**1. THE HIERARCHICAL MODEL OF LANGUAGE DOCUMENTATION.** The hierarchical model of language documentation sees as its primary goal the compilation of a LDC (Himmelmann 2006) which can function as a resource for writing descriptive grammars, dictionaries, or educational materials, as shown in Figure 1.

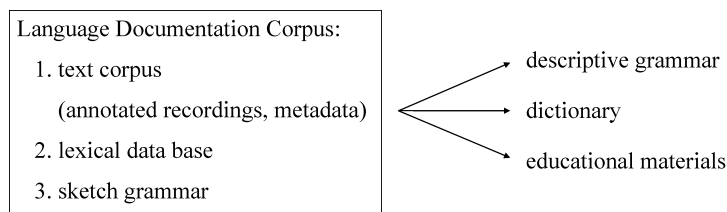


FIGURE 1: The hierarchical model of language documentation

While the exclusion of descriptive grammaticography from documentary linguistics has been criticized by several linguists (Evans 2008, Woodbury 2011), the production of educational materials does not seem to be a topic in linguistic debates, although most linguists support collaborative fieldwork and feel obliged to “give something back” to the speech community. Dobrin et al. (2009: 43) criticize the “commodification of endangered languages” as follows: “Linguists’ professional obligations to field communities are often formulated in terms of transacted objects rather than through knowledge sharing, joint

engagement in language maintenance activities or other kinds of interactionally-defined achievements.” Bown (2011: 468) remarks, “Community members often also report sometimes feeling that the linguist comes in, reifies the language, turns it into a commodity, and then takes it away.”

The problem with the hierarchical model is that even if a standard orthography has already been developed (Seifart 2006, Lüpke 2011), creating educational materials from a LDC may turn into a daunting task for a number of reasons:

1. the texts of the LDC represent spontaneous or elicited speech and may be difficult to understand because of repetitions, hesitation phenomena, and missing information
2. the texts may contain expressions like loan or swear words that are not acceptable for teachers
3. the people who are involved in the production of the EMs do not know how to exploit the LDC (Nathan & Fang 2009: 136)
4. the texts may not include contents or registers and genres that are considered useful or suitable for the prospective users of the planned materials

The first and the second problem can be solved by editing the transcripts, and the third one by training, though once the documentation is finished, it may be difficult for the linguist to find the time and the money to conduct training courses or help individual community members to derive EMs from the LDC. The fourth problem cannot be solved because the LDC simply does not provide any suitable resources for EMs.

**2. THE INTEGRATIVE MODEL OF LANGUAGE DOCUMENTATION.** The integrative model applies collaborative methods of language documentation which combine “field-work with teaching, training, and mentoring native speakers for sustainable documentation projects.” (Grinevald 2003: 60). In this approach it is the indigenous language documenters who decide on the content, the purpose, and the format of the LDC, while the linguist works as their adviser in technological, organizational, and linguistic matters, explaining to them what can be done with the available resources and which kind of genres and topics would be the most suitable to begin with (Mosel 2006).

The first products of language documentation in this model are certainly not a comprehensive dictionary and recordings, transcriptions, and edited versions of conversations, but stories that are frequently told in the community or descriptions of certain activities that are considered as useful texts and are easily recorded and transcribed. Thus, from the very beginning, the team will work on building a LDC that includes educational materials for children, teachers, or the general public, but probably excludes a collection of texts that have nothing to do with the community’s culture, such as translations from the contact language or elicited stories like the famous frog or pear stories (Chafe 1980, Mayer 1969, Chelliah & Reuse 2011: 427). In contrast to narratives and procedural texts, the recordings of natural conversations are too difficult to annotate and edit in the beginning of a LD project so that the production of conversational texts for learning-oriented materials can only be considered for a later stage of the project.

Linguists who depend on money from universities and scientific funding agencies may wonder if this kind of collaborative fieldwork is compatible with their professional aims and obligations. The answer is definitely yes if the compilation and annotation of the LDC meets scientific standards and provides a reliable basis for linguistic and other research (see Section 4).

### **3. EDUCATIONAL MATERIALS IN THE TEOP LANGUAGE DOCUMENTATION PROJECT.**

Teop is an Oceanic Meso-Melanesian language of the North-West Solomonic linkage (Lynch et al. 2002: 101 ff.), spoken by approximately 6,000 people in the Autonomous Region of Bougainville, Papua New Guinea. The project began in 2000 just after the civil war, the so-called Bougainville Crisis (1988–1999). During the first phase of the project (2000–2003), Ruth Saovana Spriggs, a native speaker of Teop, recorded interviews with elders of the community about customs and their personal histories, and I have been collecting legends, procedural texts, and encyclopedic descriptions with a team of highly motivated and skilled local language documenters since 2003. Currently (15 December, 2011), the Teop LDC consists of 258,957 words. The first book we published was a book of Teop legends (Magum, Enoch Horai, Joyce Maion, Jubilie Kamai, Ondria Tavagaga, with Ulrike Mosel & Yvonne Thiesen (eds.) 2007) because the legends seemed to present interesting reading materials for children and were easily recorded, transcribed, and edited. Not counting a hymn book printed in the 1950s, this is the very first book in the Teop language. It contains 40 edited versions of legends which were originally oral and narrated by 24 speakers, and it was produced in the following way:

1. recordings done by the linguist and native speakers
2. transcriptions done by native speakers, checked by the linguist, discussed with native speakers, revised and adjusted to a practical orthography developed by teachers in the 1980s, and eventually translated into English by the linguist with the help of native speakers
3. editorial work done by native speakers, checked by the linguist, discussed with the editor, revised and translated by the linguist with the help of native speakers; all writings done by hand because of the lack of electricity
4. oral and edited versions typed in Germany
5. proofreading of all legends, and minor changes made independently by two teachers; changes discussed with both teachers
6. preliminary version typed in Germany
7. final proofreading done in PNG
8. printing done in Germany because there was no printing press in the Autonomous Region of Bougainville
9. copies of the book sent to Bougainville and officially launched
10. all legends archived in PDF format in the DoBeS archive and made accessible without registration (Magum, Enoch Horai, Joyce Maion, Jubilie Kamai, Ondria Tavagaga, with Ulrike Mosel & Yvonne Thiesen (eds.) 2007)

Furthermore, the legends book was read by a native speaker and recorded. The recordings of her readings were annotated in ELAN with transcriptions and translations into English and were also archived in the DoBeS archive where they are freely accessible without registration (Magum, Enoch Horai, Joyce Maion, Jubilie Kamai, Ondria Tavagaga, with Ulrike Mosel & Yvonne Thiesen (eds.) 2007).

The editors and teachers were advised not to imitate the style of English stories but to keep as closely as possible to the original text, only remove hesitation phenomena and speech errors, and only make additions where absolutely necessary for the reader.

Editing the autobiographical narratives collected earlier by Ruth Saovana Spriggs was much more difficult than editing the legends. While the narration of legends followed a fixed story line with more or less conventionalized ways of expression, the personal narratives were embedded in casual conversations so that the editors felt that they had to make more changes to turn the transcriptions into readable texts. This is also reflected in word counts. The length of the edited legends is 98% of the length of the transcripts and, thus, almost the same, whereas the lengths of the edited autobiographical narratives is reduced to 42% of the transcripts.

When we began collecting procedural texts and encyclopedic descriptions for small specialized dictionaries on house-building, body and health, fishing, animals, and plants (Mahaka, Mark, Enoch Horai Magum, Joyce Maion, Naphtali Maion, Ruth Siimaa Rigamu, Ruth Saovana Spriggs & Jeremiah Vaabero, with Ulrike Mosel, Marcia Schwartz & Yvonne Thiesen 2010, Mosel 2011), most team members decided to write the texts straightaway without doing recordings and transcriptions. Having done the editorial work over several fieldwork sessions, they had become confident about their writing skills. All books are richly illustrated with drawings by an indigenous artist and photos which I produced, and will be published both in print and online in 2012 and 2013.

**4. THE TEOP LDC AS A RESOURCE FOR LINGUISTIC ANALYSIS.** The edited versions of the legends and autobiographical narratives as well as the written procedural texts and encyclopedic descriptions certainly do not represent traditional genres and indigenous registers of everyday communication. But as the data come from seven different editors and writers and have been reviewed by accepted language experts of the community, we are sure that we produced reliable language data.

An analysis of the differences between the oral legends and their edited versions showed five types of changes (Mosel 2008, forthcoming):

1. purification by the replacement of loan words
2. elaboration by the addition of words, phrases, and clauses
3. linkage of paratactic clauses by explicit coordination and embedding constructions and interlacing by raising constructions
4. compression by putting more information into a single linguistic unit, resulting in more complex structures
5. decompression by the resolution of complex structures

But the edited versions did not contain any constructions that were not found in the original transcripts.

For lexical and grammatical studies, the comparable subcorpora of oral and edited narratives show alternative ways of expressing the same content and, thus, give a fuller picture of the expressive potential of the language than the transcripts would have done by themselves. Furthermore they provide an innovative type of data for the study of the differences between spoken and written language. While the research on spoken vs. written European language varieties usually takes the structure of written varieties as the point of departure and explores how the spoken language deviates from the written one, the Teop data allow one to take the opposite perspective and study what people actually do when they put spoken texts into writing and thus develop a written language variety.

**5. CONCLUDING REMARKS.** This paper proposes a new perspective on the role of educational material in language documentation projects and describes how such resources can be utilized by community members as well as for linguistic research. Such community utilization can be further encouraged by workshops and other pedagogical means.

Due to the special circumstances after the civil war, the Teop language project started as a grass-roots project. We did not involve any official authorities but only worked together with teachers of the local village school without informing the wider public or bringing in further experts of documentary linguistics. But for the future we plan to conduct workshops for teachers and language activists (see Florey & Himmelmann 2009 for training strategies) and broaden our view of language documentation by learning from the related disciplines of language pedagogy and language revitalization (Hinton 2011, Nathan & Fang 2009).

#### REFERENCES

- Bowern, Claire. 2011. Planning a language-documentation project. In Peter K. Austin & Julia Salabank (eds.), *The Cambridge Handbook of Endangered Languages*, 459–482. Cambridge: Cambridge University Press.
- Chafe, Wallace L. (ed.). 1980. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chelliah, Shobhana L. & Willem J. de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.
- Dobrin, Lise M., Peter K. Austin & David Nathan. 2009. Dying to be counted: the commodification of endangered languages in documentary linguistics. In Peter K. Austin (ed.), *Language Documentation and Description* 6, 37–52. London: School of Oriental and African Studies.
- Evans, Nicholas. 2008. Review of essentials of language documentation. *Language Documentation and Conservation* 2(2), 340–350.
- Florey, Margaret & Nikolaus P. Himmelmann. 2009. New directions in field linguistics: Training strategies for language documentation in Indonesia. In Margaret Florey (ed.), *Endangered languages of Austronesia*, 121–140. Oxford: Oxford University Press.

- Grinevald, Colette. 2003. Speakers and documentation of endangered languages. In Peter K. Austin (ed.), *Language Documentation and Description 1*, 52–72. London: School of Oriental and African Studies.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 1–30. Berlin: Mouton de Gruyter.
- Hinton, Leanne. 2011. Language revitalization and language pedagogy: New teaching and learning strategies. *Language and Education* 25(4). 307–318.
- Lüpke, Friederike. 2011. Orthography development. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 312–336. Cambridge: Cambridge University Press.
- Lynch, John, Malcolm Ross & Terry Crowley. 2002. *The Oceanic Languages*. Richmond: Curzon Press.
- Magum, Enoch Horai, Joyce Maion, Jubilie Kamai, Ondria Tavagaga, with Ulrike Mosel & Yvonne Thiesen (eds.). 2007. *Amaa vahutate vaa Teapu. Teop legends. Illustrated by Rodney Rasin*. Kiel: Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Christian Albrechts Universität. <http://www.mpi.nl/dobes/projects/teop>.
- Mahaka, Mark, Enoch Horai Magum, Joyce Maion, Naphtali Maion, Ruth Siimaa Rigamu, Ruth Saovana Spriggs & Jeremiah Vaabero, with Ulrike Mosel, Marcia Schwartz & Yvonne Thiesen. 2010. *A inu. The Teop-English dictionary of house building*. Kiel: Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Christian Albrechts Universität. <http://www.mpi.nl/dobes/projects/teop>.
- Mayer, Mercer. 1969. *Frog, Where Are You?* New York: Penguin.
- Mosel, Ulrike. 2006. Fieldwork and community language work. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 67–85. Berlin: Mouton de Gruyter.
- Mosel, Ulrike. 2008. Putting oral narratives into writing – experiences from a language documentation project in Bougainville, Papua New Guinea. Paper given at Simposio Internacional Contacto de lenguas y documentación. Buenos Aires, 14–15 August 2008. <http://www.linguistik.uni-kiel.de/Oral%20narratives%20into%20writing.pdf> (21 March, 2012).
- Mosel, Ulrike. 2011. Lexicography in endangered languages communities. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 337–339. Cambridge: Cambridge University Press.
- Mosel, Ulrike. forthcoming. Corpus linguistic and documentary approaches in writing a grammar of a previously undescribed language In: Toshihide Nakayama & Keren Rice (eds.), *Keren Practical and Methodological Issues in Grammar Writing*.
- Mosel, Ulrike, Jessika Reinig, Marcia Schwartz, Ruth Saovana Spriggs & Yvonne Thiesen. 2007. *Das Teop Language Corpus*. <http://www.mpi.nl/dobes/projects/teop>.
- Nathan, David & Meili Fang. 2009. Language documentation and pedagogy for endangered languages: A mutual revitalisation. In Peter K. Austin (ed.), *Language Documentation and Description 6*, 132–160. London: School of Oriental and African Studies.
- Seifart, Frank. 2006. Orthography Development. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 275–299. Berlin: Mouton de Gruyter.

Woodbury, Anthony C. 2011. Language Documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 159–186. Cambridge: Cambridge University Press.

Ulrike Mosel  
umose1@gmx.de