Empowerment and Capacity Building in Endangered Speech Communities: an Example from French Polynesia

Gabriele Cablitz, Fasan Chong and Edgar Tetahiotupac

Seminar für Allgemeine und Vergleichende Sprachwissenschaft der CAU zu Kiel, Olshausenstrasse 40, 24098 Kiel, Germany

Cultural Association Te Reo o te Tuamotu, B.P. 4025, Tahiti, 98714 Papeete, French Polynesia

Laboratoire de Recherche en Sciences Humaines de Polynésie française (LARSH), Servitude Pékin, Taunoa, 98714 Papeete, French Polynesia
gcablitz@googlemail.com

ABSTRACT

This paper reports on how a team of researchers (linguistics, anthropology) within the DoBeS-program of endangered languages (Volkswagen foundation) documents and save-guards the endangered indigenous languages of the Marquesan and Tuamotuan archipelagos together with members of the speech communities. The documentation projects of the Marquesan and Tuamotuan languages are trying to capture the intimate link between language and cultural knowledge by documenting a wide range of different topics such as traditional narratives, songs, secret languages, toponymy as well as traditional practices such as food and plant medicine preparation, handicrafts, etc. by video-taping these procedures and activities and letting speakers give verbal explanations about them. The heart and success of these documentations are largely depend on the participation and engagement of speech community members. Our language documentation project has made increasing efforts to promote language documentation skills within the speech communities to ensure that language documentation efforts will continue after the short-term documentation projects of three to five years. Skills of using modern recording technologies, work with linguistic software or learning how to edit spoken texts for the development of pedagogical material have been some of the foci in the cooperation with the speech communities during fieldwork periods. This paper gives details about the cooperation with the speech communities, the linguistic software used for language documentation and the built-up of an electronic language archive.

Keywords

Endangered languages; language documentation; capacity building; linguistic fieldwork

1. INTRODUCTION

According to UNESCO estimates over half of the 6700 languages spoken world-wide will vanish in the next 100 years, and with them unique cultural knowledge, expressions of thought and social organisation. Languages are invariably linked to the culture of their speakers and “they constitute a determining factor in the identity of groups and individuals”, an increasingly important aspect in view of the on-going globalisation. UNESCO experts estimate that one language disappears approximately every two weeks and that 96 % of the world’s languages are only spoken by 4 % of the world population.1 Since 2000 the Volkswagen foundation has funded over 45 endangered language projects world-wide to prevent this irretrievable loss by financially supporting researchers from various disciplines to make audio- and video-recordings, still images, subsequent annotations (e.g. transcriptions, translations) as well as build-up an electronic multimedia archive of endangered languages. This paper reports on how a team of researchers (linguistics, anthropology) within the DoBeS-program2 of the Volkswagen foundation documents and save-guards the endangered indigenous languages of the Marquesan and Tuamotuan archipelagos together with members of the speech communities.

2. THE DoBeS-PROGRAM

When the Volkswagen foundation started its DoBeS initiative neither national or international cultural autorities nor academic disciplines had developed the capacity to ensure a comprehensive documentation of endangered languages. The program started with an initial pilot phase of one year in which eight documentation teams worked out a logistical, technical, linguistic and juridical framework of the archive on endangered languages. Together with the technical team of the Max Planck Institute for Psycholinguistics (Nijmegen, Netherlands) all teams of the pilot phase agreed on an archive with linguistic and technical standards for all documented languages in the DoBeS-program. Other important issues such as ethical and juridical questions (e.g. copyright) with detailed recommendations, constraints and contracts for all participants and potential users3 were also worked out in detail. The electronic archive is housed and maintained by the technical team of the Max Planck Institute for Psycholinguistics which also develops the linguistic software as well as the technology of maintaining and building up the archive. The Volkswagen foundation hopes to contribute with its DoBeS initiative towards the save-guarding of linguistic and cultural

1 Cf. UNESCO’s Intangible Cultural Heritage (ICH) website: http://www.unesco.org/culture/ich.
3 Details can be viewed on the DoBeS-website: http://www.mpi.nl/DOBES/ethical_legal_aspects/.
diversity by putting a particular focus on systematic language documentation. Three aspects are central in the initiative: data orientation, multifunctionality and general accessibility. Unlike other linguistic programs the DoBeS-initiative does not focus on grammatical descriptions and dictionaries, but on a wide range of primary data and speech genres (narratives, songs, procedural texts, political or ritual discourse, etc.) which should document how the language is used in various different social and cultural interactions. The minimal requirement of a documentation is a transcription and subsequent annotation (translation and explanatory comment of the speech situation and content) of primary data recordings. These language documentations should not only serve as a basis for linguistic research, but also for other disciplines (e.g. anthropology, sociology, oral history and literature, etc.), and thus the data have to be presented in an intertheoretical format. A third major concern of the Volkswagen foundation is the general accessibility and longterm usage and storage of the data. The needs and desires of the speech communities have become a major focus in the program. The documentation should ideally be used to develop pedagogical material for schools and the speech communities should get print out and CD ROM versions of the data after the project.

3. DoBeS LANGUAGE DOCUMENTATION PROJECTS ON THE MARQUESAS AND TUAMOTU ISLANDS

The documentation which was undertaken on the Marquesas and Tuamotuan islands consisted of two project phases. The first project was exclusively carried out on the Marquesas islands between 2003 and 2006. At the end of 2006, a second project was granted which also included the Tuamotuan languages. The second project is an interdisciplinary project in collaboration with the technical team of the MPI for Psycholinguistics (Nijmegen) in which we aim at creating a multimedia encyclopaedic lexicon for both language groups with the newly developed web-based lexicon tool LEXUS.

The Marquesan and Tuamotuan languages are the only endangered languages in Polynesia which are currently funded within the DoBeS-program. In the following we will briefly talk about language endangerment, the documentation topics and the methodology and linguistic software used in this project. Finally, we will talk about capacity building in and for the speech community inside and outside French Polynesia which has the purpose of empowering speech community members to continue the documentation work after the project as well as show ways of maintaining their own languages.

4. LANGUAGE ENDANGERMENT

In the DoBeS-program a language is classified as endangered if the speech community has given up – mainly for political and economic reasons - to speak the indigenous language in public in favour of the language of the surrounding dominant culture. This is indeed the case for most indigenous languages spoken in French Polynesia. A shift from the local languages to French in the public sector is a development which can be observed throughout French Polynesia, in particular since the installation of the atomic testing grounds on Moruroa in the 1960s which brought about a rapid westernisation of the local cultures [1]. The shift from the indigenous language to French has not only taken place in the public sector, but also at home. On the Marquesas, an increasing number of children have acquired French as their first language, the same can be observed on the Tuamotu islands. The linguistic situation is complex and shall not be further elaborated here. For further details of this development the reader shall be referred to [1], [2], [3] and [4].

The urgency of conducting a language documentation project became increasingly evident throughout the project. Around 20 old, knowledgable people of the Marquesan speech community – some of which have substantially contributed to this project – have died since 2003. The Tuamotuan community laments a similar situation. The documentation of these highly endangered languages has turned into a race against time.

5. SOME GENERAL ASPECTS OF THE DOCUMENTATION AND DOCUMENTED TOPICS

The first documentation project on the Marquesas between 2003 and 2006 was initiated by a linguist (G.Cablitz) and an anthropologist (E.Tetahiotupa) and a few local collaborators of the Marquesan speech community. At the beginning of the second project which started in 2006 the cultural association Te Reo o te Tuamotu with Fasan Chong (called ‘Jean Kape’) as the main collaborator joined the DoBeS-project, we started to conduct a similar style documentation as on the Marquesas by making audio- and video-recordings as well as photos/still images.

In order to create a comprehensive corpus of the Marquesan and Tuamotuan languages which tries to capture as many speech situations as possible, we have documented a wide range of topics such as story-telling, songs and dances, traditional food preparation, plant medicine, aspects of the material culture (e.g. plaiting, bark cloth (tapa) fabrication, tool fabrication), fishing and navigation techniques, plant products and gardening, trick languages, speeches, natural dialogue, radio broadcasting, games, interviews, spatial instructions, toponymy and culturally important places as well as contemporary events (e.g. political protest marches in 2004 etc.). This has created a large corpus of primary data of spoken language with a great variety of different speech genres. In recent years it has become a major objective in documentary linguistics to record as many different speech genres as possible as they reflect the linguistic potential and repertoire of its speakers.

The documentation process consists of three different stages: 1. data collection, i.e. making recordings, 2. data processing and analysis (transcription, translation and annotation) to create primary data, 3. final verification and archiving.

As we aim to create a digital multimedia archive, the collection of language data consisted primarily of audio- and video-recordings as well as still images/photos. If possible, we tended to make more video- than audio-recordings because they capture the speech situation in the best possible way. Some recordings only exist as audio files, mostly in situations where consultants or researchers were too timid to use a video recorder or the use of it was inappropriate.
For the documentation of preparation procedures (food, plant medicine, handicrafts, fishing, etc.) we asked the consultants to explain their actions and techniques in their own language. We mostly filmed the whole preparation procedure including the collection of material (e.g. plants, shellfish) used in the preparation forms. The audio-visual documentation of preparation procedures are accompanied by photos/still images of important steps in the preparation forms. We have additionally made photos of all things which are part of a particular preparation form (e.g. plants, fish, tools, final products, etc.).

All (speech) recordings were transcribed and translated together with native speakers. Translations of the transcripts exist in French as well as English. Apart from annotated video- and audio-sessions we are working on a trilingual general dictionary as well as on thematic glossaries which have become important foci of the documentation (e.g. food preparation, plant medicine, plants, fish/fishing, traditional artefacts, etc.).

The lexical entries in our lexical databases are mostly corpus-based, but have been added by field notes and collected wordlists of vernacular names for things of the natural environment (fish, birds, plants, shellfish, breadfruit varieties, etc.) which have mostly been identified by scientific names and corresponding standard common names in French and English (e.g. MQR. kariri, Lat. Physalis angulata, Fr. coqueret anguleux, petite tomate de Mexique, Eng. cutleaf groundcherry, wild tomato). Our lexical databases will not turn into a complete dictionary as there are well-known time constraints in short term documentation projects [5]. The thematic approach has the advantage to complete a sub-domain of the culture in a short period of time which often has motivating effects on the speech community [5].

6. CAPACITY BUILDING IN THE SPEECH COMMUNITIES

Our project has shown that the heart and success of these documents are largely depend on the participation and engagement of speech community members. One important aspect in this project is to build up capacities and promote language documentation skills within the speech communities in order to ensure that language documentation efforts will continue after the short-term documentation projects of three to five years. Skills of using modern recording technologies, work with linguistic and lexicographic software or learning how to edit spoken texts (e.g. narratives) for the development of pedagogical material have been some of the foci in the cooperation with the speech communities during fieldwork periods. Another important aspect in the cooperation with the speech communities is their involvement in the research and development of linguistic software outside their own communities by letting them come to Europe for a month-long stay and visit the respective research institutions. These Europe visits were also an opportunity to participate at one of the bi-annual DoBeS training weeks where participants of DoBeS-projects can be trained on field methodology, technological aspects (recording devices, techniques, etc.) and linguistic software used for language documentation and archiving.

At the planning and fund-raising stage of the project, the linguistic team made a proposal of documentation topics together with the anthropologist team. Although many of the originally proposed topics were documented, consultants made their own suggestions and contributed whatever they were prepared to share. We wanted to build up a positive documentation experience for the consultants and let them perform and explain activities where they felt most comfortable in. As a consequence we made a number of recordings and surprising discoveries we did not expect to document, among them a large corpus of over 160 Marquesan narratives.

For the processing of primary data we have in particular motivated younger members of the communities to learn to use linguistic software. On the Marquesas, linguistic software skills were acquired on an individual basis. In the Tuamotuan speech community, a number of little workshops, in particular with younger members of the community, were organised where the various linguistic skills (Transcriber, Toolbox, Lexique Pro, ELAN) were taught.

Transcriptions can be made with Transcriber or ELAN which can upload audio and/or video files in order to create time-aligned transcriptions. Each phrasal unit is connected with a beginning and an end time code of the digital audio or video file.

Apart from learning how to use basic tools for transcription and annotation, both speech communities have learned to use the lexicographic software tools Toolbox and Lexique Pro. Toolbox is a dictionary software developed for field linguists by SIL with which one can create conventional paper dictionaries. The MDF 4.0 lexical database type which we have used in our project allows complex lexicon structures containing a wealth of linguistic as well as encyclopaedic information such part of speech, phonological variants (doublets, triplets), homonym numbering, plural forms, glosses, reversal glosses for index lists (English and French), definitions in the vernacular language (e.g. Marquesan), French and English, dialectal usage and register, example sentences for each sense, dialectal variants, synonyms, antonyms, cross-references, scientific names, subentries, some vernacular encyclopaedic articles with French and English translations, the etymology of borrowed lexemes and scientific and anthropological book and website references. The collaborators in both speech communities have also learned to use to link photos, drawings and video clips in the interactive lexicon viewer and editor Lexique Pro which is multimedia follow-up version of Toolbox.

After training speech community members how to use these complex lexicon structures over a protracted period of time during fieldwork, the Marquesan collaborators have accomplished a breadfruit glossary of 34 breadfruit varieties independently (i.e. outside fieldwork periods). This breadfruit glossary does not only contain names of different breadfruit varieties, but also vernacular descriptions of the varieties as well as their usage in the traditional culture. Writing vernacular definitions of words as well as encyclopaedic articles about the usage of these plants in their own language was a first successful attempt to create a monolingual dictionary and thus giving an example of how to maintain their own language.

Language maintenance and language revival will be key aspects in the survival of these highly endangered Polynesian languages. While the writing of monolingual definitions and encyclopaedic

\[\text{SIL} = \text{Summer Institute for Linguistics, cf. website: http://www.sil.org.}\]
articles can be seen as an important means of language maintenance, they are also newly emerging speech genres which have not been part of the linguistic repertoire of speakers before. Making dictionaries is a western tradition, but often one of the most desired product of linguistic work by endangered speech communities. It is the linguist’s task to guide the local language documentators in such a way that they make use of their natural way of expressing meanings of words in a general comprehensive format of a dictionary entry, so that other speech community members will be able to understand these newly emerging speech genres (i.e. vernacular definitions and encyclopaedic articles). All in all, the two speech communities now work quite independently doing transcriptions, text editions and dictionary work.

As a consequence of this project, the core Marquesan collaborators have created a cultural association, called Te U’i hou o te Fenua ‘Enata with the aim of continuing the research and documentation of traditional and indigenous knowledge, the continuation of the trilingual dictionary and the edition of narratives for publication.

7. SUMMARY AND CONCLUSION
Although the DoBeS-program for endangered languages is a scientific program with an interdisciplinary orientation, it contributes to safe-guard linguistic and cultural knowledge of endangered speech communities in a substantial way. Its focus is set on the documentation of endangered spoken languages by creating large language corpora which contain a wide range of different speech genres. The DoBeS-archive does not only create an excellent basis for future linguistic and anthropological (and other) research, but it also allows current generations to maintain and revitalise their languages as well as future generations to reconstruct their ancestral languages. For our own documentation project, we have shown that speech community members can become major agents of the documentation, maintenance and revival of their languages and cultures. The empowerment of speech community members to become local language documentators does not only ensure that the documentation of highly endangered languages will continue after the short-term scientific projects, but it also intends to create an awareness among the scientific as well as speech community that capacity building, close cooperation and long-term continuation is part and parcel of the application of cultural human rights.

8. ACKNOWLEDGMENTS
We like to thank the Volkswagen foundation for the generous support of our DoBeS-grants. We are grateful for the enthusiastic support of members of the cultural associations Te U’i hou o te Fenua ‘Enata (Marquesas islands) and Te Reo o te Tuamotu (Tuamotu islands). We like to thank in particular Mathias (Taa’i) Tohetiaatua and Grégoire (Taiea) Epetuhui for sharing their profound knowledge of the Marquesan culture with us.

9. REFERENCES