Institution: University of Oxford

Unit of Assessment: 28 Modern Languages and Linguistics

Title of case study:
Preserving a linguistic heritage: Biak, an endangered Austronesian language

1. Summary of the impact

Biak (West Papua, Indonesia) is an endangered language with no previously established orthography. Dalrymple and Mofu's ESRC-supported project created the first on-line database of digital audio and video Biak texts with linguistically analysed transcriptions and translations (one of the first ever for an endangered language), making these materials available for future generations and aiding the sustainability of the language. Biak school-children can now use educational materials, including dictionaries, based on project resources. The project also trained local researchers in best practice in language documentation, enabling others to replicate these methods and empowering local communities to save their own endangered languages.

2. Underpinning research

Pioneering research, led by Prof Mary Dalrymple (Professor of Syntax) and postdoctoral research fellow, Dr Suriel Mofu, both based at the Faculty of Linguistics, Philology and Phonetics, at the University of Oxford, resulted in the first web-accessible repository of transcribed and linguistically analysed audio text of Biak, an endangered Austronesian language (50-70,000 speakers)[see section 3:1]. This database, developed in collaboration with Universitas Cenderawasih and Universitas Negeri Papua, is one of the first, not just for a lesser studied Austronesian language, but for **any** lesser studied language. Even now such databases are very rare.

Traditional methods of language documentation are paper-and-pencil based, relying on the transcription skills and analytic capabilities of the investigating linguist; traditional databases are often small, consisting of a few short transcribed texts as an appendix to a written grammar. In describing the emergence of documentary linguistics as an important subfield, Himmelmann (2006) notes several reasons for creating databases containing transcribed and analysed recordings of primary data, like Oxford's Biak database: (i) many under-documented languages face extinction in the near term, and it is essential to collect as much data as possible while speakers of the languages are still available; (ii) numerous under-documented languages have never been written and hence have no established orthography, the development of which is essential when building the database as well as for all other literary efforts; (iii) freely available databases allow for verification of theoretical and analytical claims about a language on the basis of primary data; (iv) properly organised and archived data constitute a solid basis for future research on the language. Recording and preserving a comprehensive range of primary language data is a vital component of linguists' response to the threat of extinction of more than half of the world's languages.

Like many languages of West Papua, Biak is under-researched, and the *UNESCO Atlas of the World's Languages in Danger* classifies it as "vulnerable". Only a few written texts were available before Dalrymple and Mofu's work, and no audio or video samples existed. Steinhauer (2005) provides a very brief sketch of Biak grammar, but no texts; a longer description in Indonesian, with a transcribed but unanalysed text, is given by Rumbrawer & Fautngil (2002). Two dissertations on the grammar of Biak have recently appeared: van den Heuvel (2006) includes an appendix with 4 short Biak texts (about 4,000-5,000 words), and Mofu (2009) includes appendices containing 7 annotated texts (about 40,000 words). All of these are transcribed texts with no accompanying audio or video files. In contrast, the database developed at Oxford consists of a more variable and larger sample of Biak material including oral texts in Biak, including songs, stories, jokes, monologues, and conversations, including representative samples from each Biak-speaking area (North, South, East, and West Biak including sub-dialects, Numfor Island, East Biak Highland, and Sor). The project collected 64 spoken texts, which are all available on the project website in the form of digitised audio or video files. For 23 of these texts, they provide transcriptions (a total of 12358 words), translations into English and Indonesian, and morphological analyses of the words in the texts, as well as XML-tagged text in computer-readable and computer-searchable form. They also provide a glossary of all words and affixes for the entire text collection.

These texts provide the basis for current and future research into Biak phonology, syntax, and

semantics, and they help to address research questions such as dialect variation. The data demonstrate conclusively that there are no morphological or syntactic differences among varieties of Biak, though there is a great deal of phonological variation. The audio database permits cataloguing of this variation, and will form the basis of efforts to reconstruct earlier stages of the language and its relation to other Austronesian languages of the region.

Based on the database, their research has covered the following issues:

- Biak nominal complementation can be marked in various ways on the clause final auxiliary verb, and previous work on this construction had not managed to untangle the semantic differences among the various complementation types. This work provides a taxonomy of complementation types and a clear explanation of their syntax and semantics. **[2]**

- Biak marks inalienable possession (involving, for example, body parts or family members) by means of morphological marking on the head noun, and alienable possession (involving acquired items) by means of marking for both the possessor and the possessed item on the determiner. These different morphological realisations of the possession relation are shown to be very similar at an abstract syntactic level. This work throws light on the syntax of possessive marking and how it relates to meaning differences involving possession. **[3]**

- Our texts have formed the basis of research on the semantics of number in Biak **[4]**. Biak distinguishes 4 numbers (singular, dual, paucal, plural). Our texts reveal that the dual and paucal numbers always have a definite interpretation, and that the paucal refers to three or more entities, while the use of the plural requires reference to at least four entities. This contradicts claims in the theoretical literature on plurality that the plural form can always be used to refer to two or more entities. This research adds to Dalrymple's focus on semantics of number in other languages such as Indonesian **[5]**, leading to another Leverhulme Research Fellowship granted to her in 2012-13.

Overall, the success of this project led to the award of Leverhulme funding to Dalrymple (2010-11) for the documentation of Dusner, another endangered language of Papua.

3. References to the research (indicative maximum of six references)

[1] Dalrymple, M. & S. Mofu "On-line language documentation for Biak (Austronesian)" http://biak.clp.ox.ac.uk/

[2] Mofu, S. 2012. Nominal clause constructions in Biak. *Linguistik Indonesia* 30(1). [Peer reviewed.] Available on request.

[3] Mofu, S. 2012. Inalienability in the Biak Language. Twelfth International Conference on Austronesian Linguistics (12ICAL), Bali, Indonesia. [Peer reviewed.] Available on request.

[4] Dalrymple, M. & S. Mofu. Semantics of number in Biak. Submitted to *Language and Linguistics in Melanesia*. Available on request.

[5] Dalrymple, M. & S. Mofu. 2012 "Plural semantics, reduplication and numeral modification in Indonesian. *Journal of Semantics*, 1-32. [Peer reviewed.] doi: 10.1093/jos/ffr015

Details of relevant grants:

ESRC Project RES-000-22-3788, "On-line language documentation for Biak (Austronesian)" (12 months, 2009-2010, PI Mary Dalrymple, Co-I Suriel Mofu, £78,858) http://biak.clp.ox.ac.uk/

Leverhulme Trust Project F/10 192/A, "Multimodal language documentation for Dusner, an endangered language of Papua", **(**2010-11, PI Mary Dalrymple, Co-I Suriel Mofu, £136,233) http://dusner.clp.ox.ac.uk/

4. Details of the impact

Biak is a vulnerable language with no firmly established orthography. Dalrymple and Mofu's research in this area is having considerable impact on preventing the language from extinction and allowing audio and written material to become available to teach future generations. As with any living language, without children speaking and writing the language, Biak would never survive. New

bilingual dictionaries (Indonesian-Biak) written using the orthography developed at Oxford will allow these children to remain truly bilingual. This not only supports the growth of educational material for teaching children, but also forms the basis for adult literacy and development of dictionaries allowing adults to develop the use of Biak in regular interaction as well as underpin development of literary material. Further details of impact are listed below.

Enabling the teaching of Biak to school children through the development of educational materials:

As the UNESCO publication "First Language First" (2005) observes, the Indonesian government advocates use of elementary students' first language in the first three years of elementary school, accompanied by instruction in Indonesian, for students who do not speak Indonesian as a first language. Thus, the Biak dictionary is crucial for use of the language among school children. The situation in Papua is difficult, however, since the population is relatively small (about 3.5 million, about 14% of Indonesia's population of 237 million), while almost 300 of the 700 languages spoken in Indonesia are spoken in Papua. This makes the task of creating local curricular material difficult, and due to these resource constraints it has been impossible to produce suitable resources for many languages, despite government policy. Consequently, resources developed at Oxford have assisted in the creation of these resources for Biak in several ways. First, the Biak orthography used in our database files has become standard, and is used in Biak-Indonesian dictionaries in schools in Biak-speaking areas of West Papua. Second, textbooks and other curricular materials are being created on the basis of our texts by Mr. Rumbrawer of UNCEN. *"Now, our children and our children's children can hear and read our language online and they can also read printed materials that your team has developed both at UNCEN Jayapura and UNIPA Manokwari... We are happy to see that many schools in Biak now use materials produced by Mr Frans Rumbrawer, one member of your team, and his colleagues in Jayapura. We hope that the regional government would provide funding so that Biak text books and dictionaries created by UNCEN and UNIPA would be produced for all schools in the Biak regency."* [see §5.1]

Providing best practice in language documentation enabling others to replicate these methods for Biak and other endangered languages:

Dalrymple and Mofu's database is valuable not only as a record of the Biak language, but as an example of best practices in language documentation for other projects to follow. The beneficiaries include the international linguistic community as well as other interested bodies in attempting to document endangered languages or indeed any language, e.g. Biak has its own page on the website of OLAC, "the Open Language Archives Community, is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (a) developing consensus on best current practice for the digital archiving of language resources, and (b) developing a network of interoperating repositories and services for housing and accessing such resources." [i] Links have also been made to their work from the West Papuan blog Selamatkan Bahasa Leluhur Kita (Save Our Ancestors' Language) [ii], and from Wilco van den Heuvel's web pages "The Biak Language in its cultural context" [iii]. There is also a page of Biak resources on DBpedia, a crowd-sourced community project which extracts structured information from Wikipedia and makes it available on the Web. The project was started and is administered by research groups from Universität Leipzig, Freie Universität Berlin, and OpenLink Software [iv].

Dalrymple and Mofu provided training in state-of-the-art language documentation practices to researchers and faculty members at University of West Papua (UNIPA) and Cenderawasih University, Papua Province (UNCEN), enabling local communities to pursue documentation efforts for the hundreds of under-documented and undocumented languages of the area. The research has formed the basis of theoretical and applied linguistic work on Biak at UNCEN and UNIPA, including a trilingual Biak-Indonesian-English dictionary [v] created using the wordlist from our project, and illustrated with example sentences from our texts and their translations into English and Indonesian; work on the dictionary is complete, and a publisher is being sought. *"I have got copies of the on-line materials and also Biak – English and Biak – Indonesian dictionaries which were produced by UNIPA as a result of your project. The dictionaries are, for me, the first comprehensive dictionaries because they contain real data and real examples from formal to informal uses of our language."* [2]

Empowering local communities to save their own endangered languages through provision of language documentation skills training:

The project's work was conducted in collaboration with Mr. Frans Rumbrawer (UNCEN) and Mr. Alfons Arsai (UNIPA), with assistance from Ms. Jinni Makabori and Ms. Denise Mambrasar, also of UNIPA, and the participation of UNIPA and UNCEN linguistics departments. The training in language documentation which Oxford provided at UNIPA bore fruit in our 2010-11 Leverhulme-funded documentation project for Dusner, a severely endangered Austronesian language of West Papua. Ms. Mambrasar and Ms. Makobari were central participants in the Dusner project, and they have recently secured scholarships for training in academic English (April-October 2013), in preparation for attending graduate school in linguistics. Mr. Rumbrawer is currently pursuing a PhD in Linguistics at a university in Central Java. Local community members have been thus empowered with skills to begin documenting other severely endangered languages of Papua.

The Biak community has been very supportive and appreciative of the results in allowing them to preserve their endangered language: "*As a young man from Biak and one of the Biak Customary Council members I felt so grateful that, finally, you made our dream comes true. Seeing our language documented online and is freely available for all Biak speakers is a real blessing.*" [2] The Chief of the Biak Customary Council writes enthusiastically about Prof Dalrymple and Dr Mofu's work on Biak, saying "*I have seen the impact of the work they have done that have motivated and strengthened several efforts by local communities to preserve and develop the language*". [3] Dr. Mofu gave a presentation of the project database in a general public meeting of the Biak community in Manokwari, West Papua (19 December 2012); the excitement and enthusiasm for the work that was generated by this meeting resulted in a follow-up prayer meeting (Manokwari, 3 February 2013), organised by members of the Biak community, to give thanks for the project and the database it produced. Each meeting was attended by about 50 Biak speakers from the Manokwari area. *"The fear that our language will disappear is now replaced with happiness that there is a hope for its survival..."* [1]

5. Sources to corroborate the impact (indicative maximum of 10 references)

Testimonial evidence:

[1] Written statement from a group of elders in the Biak community.

[2] Email statement from a Member of Biak Customary Council.

[3] Written statement from Chief of the Biak Customary Council.

Other sources of corroboration:

[i] the Open Language Archives Community (OLAC: http://www.language-archives.org/language/bhw), an on-line library of language resources about Biak;

[ii] the West Papuan blog Selamatkan Bahasa Leluhur Kita (Save Our Ancestors' Language: http://selamatkanbahasaleluhurkita.blogspot.sg/);

[iii] Wilco van den Heuvel's web pages "The Biak Language in its cultural context" (http://www2.let.vu.nl/oz/biak/links.php),

[iv] DBPedia, a collection of structured databases (http://dbpedia.org/page/Biak_language)

[v] Mofu, Suriel. 2013. *Biak-English-Indonesian Dictionary*, 98 pp. To appear.