1

Introduction: Documenting and Revitalizing AustronesianLanguages

Margaret Florey

Monash University

This chapter provides an overview of the issues and themes which emerge throughout this book. It begins with a brief description of language revitalization activities which are taking place in the Pazeh, Kahabu and Thao aboriginal communities in the mountains and plains of Taiwan. The activities of elders in these communities exemplify the growth of language activism. These case studies lead to a discussion of changes in the field of linguistics and the alliances which are being built between linguists and community language activists. The 11 chapters in the book are then reviewed within the key themes of international capacity building initiatives, documentation and revitalization activities, and computational methods and tools for language documentation.

1. LANGUAGE ACTIVISM AND PARTICIPATORY PRACTICE. In June 2007, a small group of scholars from Australia, Indonesia, the Philippines, Thailand, the UK, and the USA traveled with Taiwanese scholars and language activists to the Pazeh, Kahabu and Thao aboriginal communities in the mountains and plains of Taiwan. Led by ecologist and activist Yih-Ren (Oliver) Lin and several of his graduate students, the participants observed and learned about the range of activities which have been developed in these communities to revitalize their severely endangered languages.

Pazeh people face the challenge that their endangered plains language is not officially recognized by the Taiwanese Government. While struggling for recognition, the church community is at the center of language activities. Spurred on by linguistic assessments that only one very elderly fluent speaker of this language remains and that the language would disappear by the end of the twentieth century (Gordon 2005; Li 2000: 89), church members began supporting their language.

Presbyterian minister Rev. Daxawan Lai has become a leader in language revitalization efforts. He built a museum to collect and display Pazeh artifacts, and to provide a community focus for language and cultural activities. In 2006 a workshop was held at the museum to teach and discuss Pazeh history using the artifacts and old documents. With no academic assistance, church members have produced language learning materials and begun teaching Pazeh in the church after the Sunday service (see Figures 1 and 2). Some 400 people are said to now be learning the language.

The Kahabu (Kaxabu) people are also struggling for recognition within Taiwan and more widely. Their language was listed as a dialect of Pazeh in Ethnologue 14 (Grimes 2000) but was omitted from Ethnologue 15 with the note that "Between the 14th and 15th editions this language code was retired from use. Reason: The speech variety denoted by the code was merged into another language" (Gordon 2005).

FIGURE 1. Pazeh elders demonstrate a language lesson in their church

Although uncertain of its official future, Kahabu elders are now teaching the language to younger people in the community, and are writing new songs in Kahabu to support their language lessons (see Figure 3).

A small Kahabu museum has been built to house artifacts donated from the collections of local people (Figure 4). Both the museum and the language classes highlight the determination of the community to fight for recognition of their history and identity.

FIGURE 2. Pazeh teaching materials in the Sunday School

FIGURE 3. Kahabu elders gather for the meeting

Unlike the Pazeh and Kahabu plains tribes, Thao is one of the thirteen officially recognized indigenous tribes of Taiwan. Having achieved recognition in 2001, its status opens the way for government resources to support this very seriously endangered mountain language. Yet the community at Sun Moon Lake in central Taiwan is not waiting for outside assistance and a Thao Language Revival Committee has been formed. Thao language

Figure 4. Rev. Lai introduces Kahabu museum director Ms. Mei-Ying Huang

activism is also being supported by local anthropological scholar Mr. Shi-Lang Jen, who has spent ten years undertaking language revitalization work. Mr. Jen is currently developing an orthography and preparing pedagogical materials from Blust's (2003) Thao-English dictionary (Figures 5 and 6).

FIGURE 5. Local anthropological scholar Mr. Shi-Lang Jen discussing Thao language work with Yih-Ren (Oliver) Lin

In some aspects, these brief stories of community action in response to concern for endangered languages are not unfamiliar to linguists working with communities of speakers around the world. What is striking, however, is the determination of individuals and

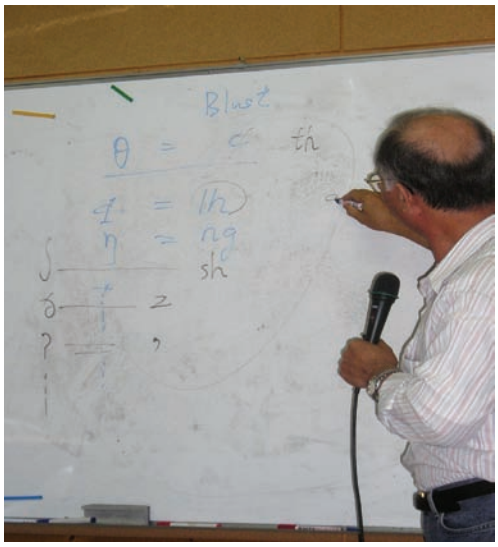

FIGURE 6. Mr. Jen illustrates the orthography he is developing from Robert Blust's (2003) Thao dictionary

groups of people within the tribes to vigorously support these languages and not let them die while the communities wait for specialist support. Although Thao and Pazeh have been the subject of quite intensive descriptive linguistic work, particularly by Paul Li, Shigeru Tsuchida and Robert Blust (see, for example, Li 2000, 1976; Blust 2003, 1996, 1999; Li and Tsuchida 2002, 2001), the linguistic publications remain largely inaccessible to community members. It was inspiring to see the materials and activities which have been produced for the revitalization programs by untrained and highly motivated people. As Peter Austin so aptly noted in a blog discussing the field trip, "What we saw contradicts the picture painted on the Academia Sinica Institute of Linguistics website: 'The present situation of the Thao can be described as one of terminal assimilation … All but one of the known speakers was born in 1937 or earlier. Some younger Thao profess an interest in learning their own language, but have little idea of how to proceed, generally having very misguided ideas based on their primary exposure to Taiwanese, Mandarin, and the Chinese writing system. The future of the Thao language seems all but sealed'" (16 June 2007).[1] In enacting Darrell Kipp's four rules (below) for setting up language programs (2000; cited in Hinton 2002: 92), the Pazeh, Kahabu and Thao people demonstrate the growth of language activism in Taiwan and internationally:

> Never ask permission; never beg to save the language.
>
> Don't debate the issues.
>
> Be very action-oriented; just act.
>
> Show, don't tell.

Language activism is a central part of the "new linguistics", which is conceptualized as "a more participatory and politicized linguistics with alliances being built between external and internal language activists who are working towards the shared goal of documenting and supporting minority languages and cultures" (Florey forthcoming), and is characterized by profound changes to ethics, methods and practice in the field. It is in settings such as the aboriginal communities described above that the impact of those profound changes to our field might most be felt. While highlighting what can be done with few resources, members of the three communities and the visiting scholars were also cognizant of what might be achieved if the language teams had access to the appropriate linguistic and pedagogical skills, methods and technologies.

A commitment to language revitalization and participatory practice with language activists and linguists working in alliance was shared by participants at the International Conference on Austronesian Endangered Language Documentation (held at Providence University in the days preceding the field trip described above), and is evident throughout the papers in this book. The themes which emerge from the chapters, and which are shaping the documentation and revitalization of Austronesian languages, are discussed in the following sections.

---

[1] See http://www.ling.sinica.edu.tw/v7-2-review.asp-v_id=9.htm.

2. DOCUMENTING AND REVITALIZING AUSTRONESIAN LANGUAGES. The contributors to this book were asked to discuss their work with Austronesian languages within the field of modern language documentation (as expounded in Himmelmann 1998; Woodbury 2003). The question of "what is documentary linguistics?" recurs in the early chapters. Ken Rehg distinguishes between three approaches to fieldwork: (1) artifactual fieldwork, (2) traditional fieldwork, and (3) documentary fieldwork. Like Rehg, Peter Austin argues that language documentation "differs fundamentally and critically from language description". Both authors point to the "ambitious and inclusive" scope of language documentation and highlight its goal of producing materials that meet the needs of both the speech and the scientific communities. Quakenbush notes that SIL linguists historically have been involved in language description, and that SIL is responding to new developments in the field of documentary linguistics and to growing international awareness of language endangerment.

The authors in this book have all approached the goals and needs of language documentation and revitalization from a number of perspectives, ranging from capacity building initiatives at an institutional or local community level, to developing frameworks for participatory practice, and pedagogical methods and outcomes, through to software development for language documentation and archiving. The papers presented here thus demonstrate the range of activities which are taking place under the rubric of language documentation in Austronesia.

2.1 INTERNATIONAL CAPACITY BUILDING INITIATIVES. A heightened concern about language endangerment is being felt throughout the Austronesian world and beyond. An increasingly important response to that concern is the provision of specialized training to linguists and to community language activists. The papers in Part 1 illustrate three different approaches to building capacity for documentation and revitalization.

Ken Rehg poses the question of how linguists of the future will remember us in relation to what we did at a time when so many languages face extinction. "Will we be admired for having conscientiously responded to this crisis, or will we be ridiculed for having thoughtlessly ignored our evident duty?". At the University of Hawai'i, Rehg and his colleagues have responded by developing a raft of activities under the Language Documentation and Conservation Initiative. Chapter 2 describes a new MA program in Language Documentation and Conservation which is training UH students to document languages. Methods in language conservation are included in the coursework. Rehg emphasizes the role of collaborative research, not only between academic researchers from different disciplinary backgrounds, but crucially, between linguists and community members. The Language Documentation Training Center plays a major role in training students to work collaboratively. This center, created by graduate students in 2003, exemplifies the new role of language activism. It links students in linguistics with the wider student population and with the general public and, in so doing, is building " a stirring sense of camaraderie".

In the four years since the Endangered Languages Documentation Program (ELDP) commenced at the the Hans Rausing Endangered Languages Program in London, grants have supported the documentation and revitalization of some twenty Austronesian languages spoken in nine countries. An MA in Language Documentation and Description has also been established in the Endangered Languages Academic Program (ELAP) along-

side a PhD program in Field Linguistics. In Chapter 3, Peter Austin argues that language documentation requires knowledge and application of a range of skills, which are taught in the MA and PhD programs. An interesting innovation in the MA is the development of two pathways: a Field Linguistics pathway and a Language Documentation and Support pathway. Applied linguistic principles which are necessary to support community language revitalization efforts are taught in the second pathway.

Adelaar (forthcoming) highlights the role which SIL has played in the Austronesian world, suggesting that it is "by far the largest single organization involved in the study and preservation of languages". This is very apparent in Quakenbush's chapter, which points to SIL involvement in 393 Austronesian language communities. In critically reviewing the history of SIL, Quakenbush notes the long-standing practice of supporting the two aspects of what he calls "language development": that is, the production of published language resources and training in language competencies such as literacy. Combined, these activities are seen to support the maintenance and/or revitalization of indigenous languages. Chapter 4 also usefully identifies and discusses four common ingredients for successful language development and language revitalization efforts—critical resources, critical expertise, critical mass, and a critical context.

2.2 DOCUMENTATION AND REVITALIZATION ACTIVITIES. The five chapters in Part 2 demonstrate the breadth of activities which are being undertaken in the Austronesian region within a participatory framework. Common themes which are explored in these chapters include government policies and legislation concerning autonomy and land tenure, and those directly informing language activities.

I Wayan Arka analyses the increased decentralization and autonomy which have characterized post-Suharto Indonesia, and compares the impact of new legislation on two languages. Balinese is one of the few examples of a language which meets the criteria for Fishman's Stage 1 in his Graded Intergenerational Disruption Scale (GIDS) for Threatened Languages (1991: 87-111). As a relatively large and high status language, Balinese continues to flourish in an environment with strong traditional social structures and leadership, and good political and financial support. In contrast, Rongga in Manggarai, Flores Island, is a minority language amongst minority languages, and has little social and political support. The issues raised in Chapter 5 are wide-ranging and Arka's case studies exemplify the situation facing an increasing number of small ethnolinguistic groups in Indonesia. He highlights the far-reaching effects which government policies can have on the prospects of language revitalization, and demonstrates that, for those prospects to be realized, priority must be given not only to capacity building but also to strengthening organizations and reforming institutions, particularly at the local level.

Four of the papers in this section focus on language activities in Taiwan and broaden our understanding of the issues which aboriginal tribes are facing and the responses which are being developed in partnerships between community members and academic researchers. Fuhui Hsieh and Shuanfan Huang have been documenting Kavalan, a seriously endangered language spoken in southeastern Taiwan. The authors use sociolinguistic and demographic data to analyze the shift from Kavalan existing in a multilingual setting to the use of Mandarin and Taiwanese by younger Kavalan people. The recent educational policy introduced in Taiwan potentially supports aboriginal languages but Hsieh and Huang re-

port that Kavalan people have not yet been able to organize to access such support. They promote cooperation between linguists and local leaders and language activists to establish language revitalization programs. Chapter 6 also gives a useful overview of the various archiving projects which have been developed in Taiwan. The authors focus in particular on the NTU (National Taiwan University) Corpus of Formosan Languages, which currently houses spoken texts in Saisiyat, Kavalan, Amis and Tsou.

The next generation of linguists in Taiwan are being mentored in language documentation and revitalization through their involvement in the work of D. Victoria Rau and Meng-Chien Yang at Providence University. Rau and Yang have collaborated with local people on creating methods for revitalizing the Yami language of Orchid Island. In Chapter 7, they analyze the development and deployment of e-learning for Yami. A questionnaire was used to assess the interest in and likely uptake of e-learning amongst the target population and found a strong positive response. Community people were involved in the development of animations. Chapter 9 investigates a different aspect of the Yami project, focusing on the acquisition of Yami as a second language. This paper contributes both to language revitalization and to the field of second language acquisition through its analysis of the methods of teaching and learning an endangered language. Rau et al point out that to position the teaching of an endangered indigenous language as a "foreign" language, even in its own country, may increase its prestige, visibility and status and hence the desirability of learning the endangered language.

Chapter 8 brings an academic researcher, Yih-Ren (Oliver) Lin, together with two Atayal researchers and activists, Lahuy Icyeh and Da-Wei Kuan (Daya), to analyze the implementation of the Taiwan government's "New Partnership Policy". Lin et al argue that the effectiveness of the policy was compromised by lack of indigenous involvement in exercises such as the mapping of Atayal territory and the study of traditional ecological knowledge. They contend that Taiwan's indigenous languages are endangered because of alienation from their socio-political contexts, and that it is essential to provide an environment in which indigenous languages can thrive. The Atayal project in Smangus village provides a case study of the way in which communities and academic institutions can come together to develop a curriculum which simultaneously supports the revitalization and strengthening of indigenous languages and traditional ecological knowledge whilst also providing an environment in which non-indigenous people can be exposed to and learn about those practices in their homeland context.

2.3 COMPUTATIONAL METHODS AND TOOLS FOR LANGUAGE DOCUMENTATION. Computational tools are a core part of the work of language documentation and are critical in making multipurpose language data more widely accessible to a range of users, including language activists, members of the wider speech community, linguists, and educators. Programs such as Toolbox (to interlinearize, gloss and analyze data, and to build a lexicon), ELAN (to segment, time-align and transcribe audio and video files), Audacity (to capture audio data from analogue or DAT recorders), and IMDI (for detailed recording of metadata) are now widely in use. Training in these tools is not only a part of academic programs (such as those described in Part 1 of this book), but is also included in training programs designed for community language activists, such as InField (The Institute on

Field Linguistics and Language Documentation[2]) and two Indonesian Training Workshops on Language Documentation (Florey and Himmelmann forthcoming; Florey forthcoming). As linguists have become more aware of the language development needs of communities, so too have new computational tools aimed to meet those needs. The papers in Part 3 focus on the design of three tools which support language documentation and language revitalization both in the Austronesian region and more widely.

WeSay is an application which facilitates the direct involvement of language activists in dictionary compilation. Developers Albright and Hatton both have a background in software development, and share a belief in what they describe as the "evident rightness" of helping interested community members play whatever role they can in language work. In building *WeSay*, they have recognized that the tools which are commonly used in linguistic work (such as those listed above) require both extensive training and ongoing support, and thus potentially limit the number of people in a community who might be skilled in their use. The discussion in Chapter 10 demonstrates how this purpose-built program minimizes the training load and provides a user-friendly means for an individual or group of speakers to directly enter lexical data building on semantic fields.

Like Albright and Hatton, Meng-Chien Yang and colleagues at Providence University were concerned with the difficulties which language activists can face in using programs such as Toolbox and Lexique Pro. Yang et al were motivated by their experience with the Yami people of Orchid Island to develop an online Formosan Multimedia Dictionary for dictionary compilation and sharing linguistic resources. Chapter 11 describes how registered users can access this digital archive both to enter data and to search the Formosan language and dialect database.

Chapter 12 introduces the development and use of *Discourse Profiler*, a new software program created by Phil Quick which contributes to language documentation. *Discourse Profiler* is a tool for annotating discourse information in texts. Its two primary capabilities are to create a representation or a 'map' of the structure and elements of a text, and to quantify texts with an array of sixteen different possible statistical outputs. In the development phase it has been trialed on the endangered Pendau language of Sulawesi, Indonesia and has been used with Balinese data. As Quick notes, this program will be able to reveal additional and rich information that can be used for the conservation and revitalization of endangered languages.

3. ACKNOWLEDGEMENTS. This book developed from the International Conference on Austronesian Endangered Language Documentation which was held at Providence University in Taiwan from 5-7 June 2007. The conference organizing committee consisted of four members. Prof. D. Victoria Rau (Department of English Language, Literature, and Linguistics) and Prof. Meng-Chien Yang (Department of Computer and Communication Engineering) were the two major organizers at Providence University. Dr Margaret Florey (Linguistics Program, Monash University, Australia) selected the invited speakers from outside of Taiwan and worked with Prof. Rau to finalize the conference program. Prof. Yih-Ren (Oliver) Lin (Department of Ecology, Providence University) planned and led the

[2] See <http://www.linguistics.ucsb.edu/faculty/infield/>; to be held for the first time at the University of California at Santa Barbara in June-July 2008.

post-conference excursion to the Thao, Pazeh and Kahabu communities in Sun Moon Lake and Puli Township, Nantou County.

Our first thanks are to the keynote speakers Prof. Paul Li and Prof. Peter Austin, and to all of the linguists who took part in the conference and whose participation made it an acclaimed success. Participants were invited to submit their papers for possible inclusion in this volume. All submitted papers underwent a process of anonymous refereeing by two readers, and papers were selected for publication on the basis of the reviews. We are very grateful to the twenty-nine international linguists who so generously agreed to review the papers and who provided valuable feedback to the authors. We thank Ken Rehg for inviting us to produce this book and for his support and advice throughout the process, Akiemi Glenn for her efficient and dedicated editing work to finalize the book, Meng-Chien Yang for so ably producing the HTML files of this book, Hsin-Kuang Hsueh for designing the cover of the book, and Elaine Rau for additional work on the final version of the cover.

The conference was made possible by wide-ranging support both from within Taiwan and internationally, and we would like to express our deep gratitude to the following people and organizations. Dean K.-P. Tse of the College of Foreign Languages and Literature and Prof. E.-C. Wu, previous chair of the Department of English Language, Literature, and Linguistics at Providence University, offered both encouragement and substantial financial support to the organization of the conference at that university. Prof. D.-A. Ho, Chair of Institute of Linguistics at Academia Sinica, and Prof. C.-R. Huang, Chair of Linguistic Society of Taiwan, provided invaluable suggestions and agreed to co-sponsor the conference. The Council of Indigenous Peoples and National Science Council of R.O.C. (Taiwan) both responded to our requests for financial support by providing generous grants. Prof. Peter Austin of the Hans Rausing Endangered Languages Project, SOAS, University of London, gave positive feedback to the initial conference proposal and a HRELP grant to support its organization.

We would also like to gratefully acknowledge several other people who unstintingly devoted their time and talents to organizing the conference. Prof. C.-L. Luo, the incumbent chair of the Department of English Language, Literature, and Linguistics, offered administrative support by recruiting her administrative assistants and teaching assistants to join the organizing team. Teresa Hsu coordinated all the administrative work related to the preparation of the conference, while Teresa Hung, Sincere Tsai, Elli Lin, Chris Lu, and Joyce Liu all contributed their energy to handle their assigned responsibilities. Team members on Rau and Yang's ELDP project, H.-S. Guo and C.-P. Chen, helped build the conference website, while Ann Chang, Jasmine Lin, H.-T. Chou, Betsy Yang, Davis Tai, and Karen Yang joined us in co-authoring research papers for the conference. In addition, the graduate students formed a hospitality group to welcome the invited speakers. Daniel Rau (Macalester College, Minnesota, USA) served as a most capable bilingual master of ceremonies for both the conference program and banquets.

This book was completed while both editors were on sabbatical leave. Margaret Florey was a visiting scholar at the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, and Victoria Rau was a visiting scholar at the English Language Institute, University of Michigan, USA. We thank the MPI, ELI and Prof. Gunter Senft, and Prof. John Swales in particular for hosting our visits and providing stimulating and convivial work environments.

References

Adelaar, Alexander K. Forthcoming. An overview of language documentation in the Austronesian world. In *Endangered Languages of Austronesia*, edited by M. Florey. Oxford: Oxford University Press.

Austin, Peter K. 16 June 2007. Endangered languages and Taiwan. In *Transient Languages & Cultures blog*, edited by J. Simpson: http://blogs.usyd.edu.au/elac/2007/06/endangered_languages_and_taiwa.html.

Blust, Robert A. 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35(2):272-294.

Blust, Robert A. 1999. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38:321-364.

Blust, Robert A. 2003. *Thao dictionary*, *Language and linguistics Monograph Series No. A5*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Fishman, Joshua A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, UK: Multilingual Matters.

Florey, Margaret. Forthcoming. Language activism and the "new linguistics": Expanding opportunities for documenting endangered languages in Indonesia. To appear in *Language documentation and description* vol.5, edited by Peter K. Austin. London: SOAS.

Florey, Margaret, and Nikolaus P. Himmelmann. Forthcoming. New directions in field linguistics: Training strategies for language documentation in Indonesia. In *Endangered languages of Austronesia*, edited by Margaret. Florey. Oxford: Oxford University Press.

Gordon, Raymond G., Jr. 2005. Ethnologue: Languages of the World, Fifteenth edition: SIL International.

Grimes, Barbara Fox. 2000. *Ethnologue: languages of the world*. 14th ed. Dallas: Summer Institute of Linguistics (SIL).

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-195.

Hinton, Leanne. 2002. *How to keep your language alive*. Berkeley: Heydey books.

Kipp, Darrell. 2000. *Encouragement, guidance, insights, and lessons learned for native language activists developing their own tribal language programs*. St. Paul: Grotton Foundation.

Li, Paul Jen-kuei. 1976. Thao phonology. *The Bulletin of the Institute of History and Philology, Academia Sinica* 47(2):219-244.

Li, Paul Jen-kuei. 2000. Some aspects of Pazeh syntax. In *Grammatical analysis: morphology, syntax, and semantics*. *Studies in honour of Stanley Starosta*, edited by V. P. de Guzman and B. Bender. Honolulu: Oceanic Linguistics Special Publication No. 29.

Li, Paul Jen-kuei, and Shigeru Tsuchida. 2001. *Pazeh dictionary*, *Language and Linguistics Monograph Series No. A2*. Taipei: Academia Sinica.

Li, Paul Jen-kuei, and Shigeru Tsuchida.. 2002. *Pazeh texts and song*, *Language and Linguistics Monograph Series No. A2-2*. Taipei: Academia Sinica.

WOODBURY, TONY. 2003. Defining documentary linguistics. In *Language documentation and description*, edited by Peter K. Austin. London: SOAS.

Margaret Florey

margaret.florey@arts.monash.edu.au