

# Documenting and Revitalizing Austronesian Languages



*edited by*

**D. Victoria Rau and Margaret Florey**

# Documenting and Revitalizing Austronesian Languages

*edited by*

D. Victoria Rau and Margaret Florey



Language Documentation & Conservation Special Publication No. 1

PUBLISHED AS A SPECIAL PUBLICATION OF LANGUAGE DOCUMENTATION & CONSERVATION

LANGUAGE DOCUMENTATION & CONSERVATION  
Department of Linguistics, UHM  
Moore Hall 569  
1890 East-West Road  
Honolulu, Hawai'i 96822  
USA

<http://nflrc.hawaii.edu/lhc>

UNIVERSITY OF HAWAI'I PRESS  
2840 Kolowalu Street  
Honolulu, Hawai'i  
96822-1888  
USA

© All texts and images are copyright to the respective authors. 2007

© All chapters are licensed under Creative Commons  
Attribution Non-Commercial No Derivative License

*Library of Congress Cataloging in Publication data*  
ISBN 978-0-8248-3309-1



# Contents

<i>Contributors</i>	iii
1. Introduction: documenting and revitalizing Austronesian languages <i>Margaret Florey</i>	1
<b>Part I. International capacity building initiatives</b>	
2. The language documentation and conservation initiative at the University of Hawai'i at Mānoa <i>Kenneth L. Rehg</i>	13
3. Training for language documentation: Experiences at the School of Oriental and African Studies <i>Peter K. Austin</i>	25
4. SIL International and endangered Austronesian languages <i>J. Stephen Quakenbush</i>	42
<b>Part II. Documentation and revitalization activities</b>	
5. Local autonomy, local capacity building and support for minority languages: Field experiences from Indonesia <i>I Wayan Arka</i>	66
6. Documenting and revitalizing Kavalan <i>Fuhui Hsieh and Shuanfan Huang</i>	93
7. E-learning in endangered language documentation and revitalization <i>D. Victoria Rau and Meng-Chien Yang</i>	111
8. Indigenous language-informed participatory policy in Taiwan: A socio-political perspective <i>Yih-Ren Lin, Lahwy Icyeh, Da-Wei Kuan (Daya)</i>	134
9. Teaching and learning an endangered Austronesian language in Taiwan <i>D. Victoria Rau, Hui-Huan Chang, Yin-Sheng Tai, Zhen-Yi Yang, Yi-Hui Lin, Chia-Chi Yang, and Maa-Neu Dong</i>	162

### **Part III. Computational methods and tools for language documentation**

- |     |  |     |
|-----|--|-----|
| 10. | WeSay, a tool for engaging communities in dictionary building<br><i>Eric Albright and John Hatton</i>  | 189 |
| 11. | On designing the Formosan multimedia word dictionaries<br>by a participatory process<br><i>Meng-Chien Yang, Hsin-Ta Chou, Huey-Shiuan Guo,<br/>and Gia-Pyng Chen</i> | 202 |
| 12. | Annotating texts for language documentation<br>with Discourse Profiler's metatagging system<br><i>Phil Quick</i>   | 219 |

## *Contributors*

**ERIC ALBRIGHT** received his MA in Applied Linguistics from GIAL (Graduate Institute of Applied Linguistics, Texas) in 2000, and earlier, worked as a software developer. Joining SIL International, he worked among the Bidayuh of Sarawak, Malaysia from 2001-2003, advising them in the fields of literacy, orthography development and lexicography. In 2006, he co-founded Payap Language Software, part of the Linguistics Institute at Payap University in Thailand.

**I WAYAN ARKA** is a Fellow in Linguistics at the Research School of Pacific and Asian Studies (RSPAS) at The Australian National University. He received his PhD from Sydney University in 1999. He returned to Udayana University in Bali-Indonesia for three years before he moved again to Australia (in April 2001) to take up a fellowship at the RSPAS, ANU. Wayan's interests are in descriptive, theoretical and typological aspects of Austronesian languages of Indonesia. His recent books published by Pacific Linguistics are *Balinese Morphosyntax: A lexical-functional approach* (2003) and *The many faces of Austronesian voice systems: some new empirical studies*, co-edited with Malcolm Ross (2005). Wayan is currently working on a number of projects including research on voice in the Austronesian languages of eastern Indonesia (funded by an NSF grant, 2006-2009, in collaboration with Matt Shibatani and Fay Wouk), and a large-scale machine-usable grammar, dictionary and corpus of Indonesian (funded by an ARC Discovery grant, 2008-10, in collaboration with Jane Simpon, Avery Andrews and Mary Dalrymple). He has done extensive fieldwork for his Rongga Documentation Project (funded by the ELDP, 2004-6), and organised capacity building programmes for language documentation, maintenance and revitalisation in Indonesia.

**PETER K. AUSTIN** joined SOAS in October 2002 as the Märit Rausing Chair in Field Linguistics and Director of the Endangered Languages Academic Programme. Earlier, Peter held a Humboldt Prize at Frankfurt University and was Foundation Professor of Linguistics at the University of Melbourne (1996-2002). Peter's research interests cover descriptive, theoretical and applied linguistics. He has extensive fieldwork experience on Australian Aboriginal languages and since 1995 has been researching Sasak and Samawa, Austronesian languages from Lombok and Sumbawa islands, eastern Indonesia. His theoretical research is mainly on syntax and focuses on Lexical Functional Grammar, morpho-syntactic typology, theory and practice of language documentation and multi-media for endangered languages. He has also published on historical and comparative linguistics, typology, and Aboriginal history and biography.

**HUI-HUAN ANN CHANG** received her MA in linguistics from the Department of English Language, Literature, and Linguistics, Providence University. She has been a research assistant in various Yami projects since 2002, coauthoring papers and providing computer support. Her research interest is in interlanguage variation with a special focus on phonological acquisition. She has co-authored with Victoria Rau "A tale of two diphthongs in an indigenous minority language", a paper in Stanford and Preston's co-edited anthology on *Variationist Approaches to Indigenous Minority Languages* (forthcoming).

**GIA-PYNG CHEN** is a lab manager in the College of Informatics, Providence University. Her research interests are database management and web applications.

**HSIN-TA CHOU** is a graduate student majoring in Computer Sciences and Information Management at Providence University. His research interests are animation and database management.

**MAA-NEU DONG** is a board member of the Committee of Yami language certification at the Council of Indigenous Peoples in Taiwan. As a native speaker of Yami, she has published since 1989 journal articles on translations of Yami folklores and ceremonies, and a collection of Yami traditional songs “Praising on Taro” (1995). She has collaborated with Victoria Rau on Yami Textbook (2000), Yami Texts with Reference Grammar and Dictionary (2006), and papers on “Yami reduplication”, “Endangered language documentation and transmission”, and “A tale of two diphthongs in an indigenous minority language”. Her current research is on a Yami Dictionary project.

**MARGARET FLOREY** is a Senior Lecturer in the Linguistics Program at Monash University. She received her PhD in Linguistics from the University of Hawai’i at Manoa (1990). Margaret’s research interests include the minority languages of the Austronesian and Australian language families, language endangerment, ethnography, ethnobiology, and anthropological linguistics. Margaret has extensive fieldwork experience in Central Maluku, eastern Indonesia, and in Western Australia. She is currently coordinating a collaborative research project which is documenting four endangered Moluccan languages, both in the Indonesian homeland and in the Dutch diaspora. In addition to language documentation, Margaret is involved in capacity building with linguists and community language activists in Indonesia, the Netherlands and Australia. She is a co-founder of the Resource Network for Linguistic Diversity, serves on the Board of Governors of Terralingua, and chairs the steering committee for the International Conference on Austronesian Linguistics. Margaret also contributes as an international expert at the UNESCO Intangible Cultural Heritage Section Experts’ Meetings on the Convention on Intangible Cultural Heritage.

**HUEY-SHIUAN GUO** is a teaching assistant of Department of Computer Sciences and Information Management, Providence University, Taiwan. Her research interests are database management and web applications.

**JOHN HATTON** worked in Apple Computer’s Advanced Technology Group before joining SIL International in Papua New Guinea in 1997. There he served as a language software consultant and developer, working with both expatriate linguists and nationals interested in developing their own languages. He lived (and developed software) among the remote Edolo people for two years while his wife helped them start adult literacy programs. In 2006 he moved to Thailand to co-found Payap Language Software. John holds a Master’s Degree in Computer Science from the University of California, San Diego.



**FUHUI HSIEH** received her Ph.D. in linguistics from National Taiwan University. Her research interests are in syntax, semantics, and Austronesian linguistics. She has co-authored with Shuanfan Huang “Grammar, construction, and Social Action: A Study of the Qishi Construction (2005)”, published in *Language and Linguistics*, and “The Pragmatics of Case Marking in Saisiyat” (2006), published in *Oceanic Linguistics*.

**SHUANFAN HUANG** is Professor of Linguistics in the Graduate Institute of Linguistics at National Taiwan University. His particular research interests focus on the discourse basis of syntax, cognitive semantics and Austronesian linguistics.

**LAHWY ICYEH** is a member of the Smangus village of the Atayal people in Taiwan. He is currently a masters student of the Institute of Ecology, Providence University. His research interests include ethnobotany, traditional territory mapping, eco-tourism and self-governing natural resources management. He is an active member in promoting traditional territory mapping and self-governance for the Atayal people. Particularly, he is dedicated to the revitalization of Atayal people’s traditional knowledge in the modern context.

**DA-WEI (DAYA) KUAN** is a member of the Atayal people in Taiwan. He received his MA in ethnology from National Cheng-Chi University. He is currently a PhD candidate in the Department of Geography, University of Hawaii at Manoa, and also an indigenous visiting researcher (2007) in the Institute of Ethnology, Academia Sinica, Taiwan. Daya’s research interests include the relations between modernity and indigeneity, political ecology and indigenous community mapping. He has been participating in the projects of Taiwan indigenous traditional territory survey and community mapping since 2003.

**YI-HUI (JASMINE) LIN** received her MA in linguistics from the Department of English Language, Literature, and Linguistics, Providence University. Her research interest is in sociolinguistics with a special focus on language shift and the linguistic vitality of Yami.

**YIH-REN (OLIVER) LIN** received his Ph.D. in Geography from University College London. As a cultural geographer and ecologist, his research interests are traditional ecological knowledge, participatory action research methodology, and political ecology. He is actively involved in projects of revitalizing Taiwan’s indigenous peoples’ culture, which include Indigenous peoples’ traditional territory mapping and a community-based digital local knowledge database. His most recent publications include a paper (2007) on “The Chief Seattlization of Nature Conservation in Taiwan: A Radical Perspective on Indigenous Traditional Ecological Knowledge.” in *Chung Wai Literary Quarterly*. He is making a film in collaboration with local Atayal communities. This film is based on Atayal traditional ecological knowledge and will be the first one spoken in Atayal and played by community members in Taiwan.

**J. STEPHEN QUAKENBUSH** has served as a linguist/translator with the Agutaynen language community of Palawan, Philippines (1984-1997), and as Director of SIL Philippines (1997-2001). Since 2003, Steve has been Academic Services Coordinator for SIL in

Asia. His interests are in the linguistics and sociolinguistics of Agutaynen, and of Austronesian more generally. Representative publications include his 1986 Georgetown University Ph.D. dissertation *Language use and proficiency in a multilingual setting* (published in 1989 by the Linguistic Society of the Philippines) and an article entitled “Word order and discourse type: an Austronesian example” in Doris L. Payne (ed.) *Pragmatics of word order flexibility* (1992, John Benjamins).

**PHIL QUICK** completed his PhD in linguistics in 2004 at the Australian National University, and has worked in Sulawesi, Indonesia since 1987. He has a book in press with Pacific Linguistics, ‘A Grammar of the Pendau Language’, and has published about 20 articles on the Pendau language in English and Indonesian. He has been a member of SIL International since 1983, and serves as a translator, linguist and linguistics consultant for the Indonesia Branch. Phil has been working with the Pendau language team to publish vernacular materials. A book on ten Pendau folktales is also forthcoming.

**D. VICTORIA RAU** received her Ph.D. in linguistics from Cornell University. Her research interests are in Austronesian linguistics, sociolinguistics, and applied linguistics. Her most recent publications include two websites on Yami Language Documentation (2005-2007), a monograph on Yami Texts with Reference Grammar and Dictionary (2006), and “A tale of two diphthongs in an indigenous minority language”, a paper in Stanford and Preston’s co-edited anthology on *Variationist Approaches to Indigenous Minority Languages* (forthcoming). She was a visiting scholar at CARLA, University of Minnesota (2003-2004), and is currently taking her sabbatical leave from Providence University at the English Language Institute, University of Michigan.

**KENNETH L. REHG** is an Associate Professor in the Department of Linguistics at the University of Hawai‘i at Mānoa, where he serves as coordinator of the department’s activities in language documentation and conservation. His research interests include language documentation, Micronesian linguistics, phonology, lexicography, and the application of linguistics to the development of educational policies and practices in the developing nations of the Pacific. He is the founding Editor of the new on-line journal *Language Documentation and Conservation* and a member of the Linguistic Society of America’s Committee on Endangered Languages and their Preservation. He has extensive fieldwork experience in Micronesia and has served as a linguistic consultant to the departments of education in Hawai‘i, the Federated States of Micronesia, the Commonwealth of the Northern Mariana Islands, and American Samoa. He is currently working with colleagues on Pohnpei to produce a second edition of the Pohnpeian dictionary.

**YIN-SHENG (DAVIS) TAI** is an MA student in Linguistics in the Department of English Language, Literature, and Linguistics, Providence University. His research interests are in lexical semantics and Yami traditional songs.

**CHIA-CHI (BETSY) YANG** received her MA in English Language Teaching from the Department of English Language, Literature, and Linguistics, Providence University. Her research interest is in computational assisted language learning with a special focus on Yami e-Learning.

**MENG-CHIEN YANG** is an assistant professor of the Department of Computer Sciences and Communication Engineering, Providence University, Taiwan. His research interests are digital archiving, computerized tools for field linguists, and image processing.

**ZHEN-YI (KAREN) YANG** is an MA student in Linguistics in the Department of English Language, Literature, and Linguistics, Providence University. Her research interest is in sociolinguistic variation with a special focus on sociophonetics in Taiwanese languages.

# 1

## Introduction: Documenting and Revitalizing Austronesian Languages

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. Pateh 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).<sup>1</sup> 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.

---

<sup>1</sup> See [http://www.ling.sinica.edu.tw/v7-2-review.asp-v\\_id=9.htm](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<sup>2</sup>) 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

---

<sup>2</sup> 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](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](mailto:margaret.florey@arts.monash.edu.au)

# 2

## The Language Documentation and Conservation Initiative at the University of Hawai‘i at Mānoa

Kenneth L. Rehg

*University of Hawai‘i at Mānoa*

Since its inception in 1963, the Department of Linguistics at the University of Hawai‘i at Mānoa (UHM) has had a special focus on Austronesian and Asian languages. It has supported and encouraged fieldwork on these languages, and it has played a major role in the development of vernacular language education programs in Micronesia and elsewhere. In 2003, the department renewed and intensified its commitment to such work through what I shall refer to in this paper as the *Language Documentation and Conservation Initiative* (LDCI). The LDCI has three major objectives. The first is to provide high-quality training to graduate students who wish to undertake the essential task of documenting the many underdocumented and endangered languages of Asia and the Pacific. The second is to promote collaborative research efforts among linguists, native speakers of endangered and underdocumented languages, and other interested parties. The third is to facilitate the free and open exchange of ideas among all those working in this field. In this paper, I discuss each of these three objectives and the activities being conducted at UHM in support of them.

**1. INTRODUCTION.**<sup>1</sup> It is likely that linguists of the future will remember this century as a time when a major extinction event took place, as an era when thousands of languages were abandoned by their speakers in favor of languages of wider communication. What is considerably less certain, however, is how linguists of the future will remember *us*. Will we be admired for having conscientiously responded to this crisis, or will we be ridiculed for having thoughtlessly ignored our evident duty?<sup>2</sup>

Obviously, there is no way we can know the answer to this question. The future is unknowable and incalculable, but this does not mean that we cannot play a role in shaping it. As Alan Kay has asserted: “The best way to predict the future is to invent it.”<sup>3</sup> While there is admittedly a considerable amount of Western, liberal hubris implicit in this statement, it requires no unusual prescience to know that what we must do now is take action. We need to document as many endangered languages as possible, and we need to train young linguists to carry out this work.

---

<sup>1</sup> I wish to thank Peter Austin, Lisa Ebeling, Paul Newman, Laura Robinson, Tsz-him Tsui, Albert J. Schütz, the audience of the International Conference on Austronesian Endangered Language Documentation, and two anonymous referees for their useful comments on an earlier version of this paper. I am also indebted to Joel Bradshaw, with whom I have had numerous interesting conversations about fieldwork. I accept full responsibility for the shortcomings of this work.

<sup>2</sup> Michael Krauss (1992:8) noted: “If we do not act, we should be cursed by future generations for Neronically fiddling while Rome burned.”

<sup>3</sup>See <http://www.smalltalk.org/alankay.html>.



It is this latter task—the training of young linguists—that I wish to consider in this paper. Clearly, we need to train a cadre of young scholars who will conduct high-quality fieldwork in the hope that their efforts will help ameliorate the massive loss of accumulated wisdom and the catastrophic loss of information that will result if we fail to respond to this impending crisis. First, though, we must ask ourselves, how do we train them, and what do we train them to do?

This paper provides some provisional answers that we have arrived at within the Department of Linguistics at the University of Hawai'i at Mānoa (UHM)—a department that has had a long-standing interest in Austronesian languages.<sup>4</sup> Because the efforts I describe in this paper were developed within this specific context, they may be of limited applicability to linguists working elsewhere. I should also emphasize that our efforts are undergoing periodic revision, and, we hope, improvement, so this paper should be read as a report on a program under construction.

**2. LEVELS OF ADEQUACY.** If our ultimate goal is to prepare young linguists to do fieldwork, the first question we must consider is, what kind of fieldwork? What type of fieldwork will be adequate for documenting an endangered language, or any language for that matter? I would suggest that it might be useful to conceptualize this issue in terms of three approaches to fieldwork, which I will characterize as (1) artifactual fieldwork, (2) traditional fieldwork, and (3) documentary fieldwork.<sup>5</sup>

By “artifactual fieldwork”, I mean fieldwork that is done for special purposes.<sup>6</sup> It characteristically entails gathering data centered on one or more specific features of one or more languages. Fieldwork of this type is typically undertaken by comparativists, typologists, or formalists in search of information relevant to the development of a particular theoretical claim. The data, or artifacts, they gather are extracted from their natural context and are assigned significance only insofar as they are useful for the purposes of external comparison.

I use the label “traditional fieldwork” to characterize fieldwork that has as its goal the description of a specific language. The defining characteristic of fieldwork of this type is that it is generalization- rather than data-oriented. Thus, the grammars that result from such fieldwork typically include minimal amounts of data, usually just enough to illustrate a grammatical claim. While the products of such fieldwork may also include dictionaries and a limited number of texts, audio and video recordings and data bases are not usually made part of the public record.

Much of the linguistic fieldwork that has been carried out up to now can be characterized as being one of these first two types. The claims about a language that result from such work are essentially (and, until recently, necessarily) of the “take-my-word-for-it” type; they cannot be verified empirically, except through additional fieldwork. However, an increase

---

<sup>4</sup> The URL for the UHM Department of Linguistics is <http://www.ling.hawaii.edu/>.

<sup>5</sup> The term ‘fieldwork’ means different things to different people. I consider here just those types of fieldwork that focus in part or in whole on obtaining information about adult grammars of one or more languages.

<sup>6</sup> The term “artifactual fieldwork” comes from Joel Bradshaw.

in concerns about language endangerment, coupled with new technology, has given rise to a third approach to fieldwork—documentary fieldwork.

“Documentary fieldwork” is far more ambitious and inclusive than either artifactual or traditional fieldwork. Nikolaus Himmelmann (2006:1) notes that “... a language documentation is a lasting, multipurpose record of a language.” The goal of documentary fieldwork, then, is to contribute to the creation of such a record. Ideally, its outcome is a body of materials that meets the needs of both the speech and the scientific communities. To paraphrase Rhodes et al. (n.d.), it involves the development of high-quality grammatical materials and an extensive lexicon based on a full range of textual genres and registers, as well as audio and video recordings, all of which are fully annotated, of archival quality, and publicly accessible.

In reality, fieldwork is likely to combine some aspects of all of these approaches and will rarely or never achieve the high standards set for documentary fieldwork. The goals one sets for fieldwork are dependent upon many factors, including the vitality of the language, the number of people participating, the skills of the fieldworker(s), the amount of available funding, and the time available to spend in the field. It is not my intent to denigrate fieldwork of any kind. Even if we know that our efforts will fall short of the ambitious agenda of documentary fieldwork, that should not inhibit us. Indeed, it is essential that we train young fieldworkers to be pragmatic, to set realistic goals, and to assign priorities. Every piece of information that we collect is potentially useful in ways that we may not be able to envision. All fieldwork that is well-done makes a contribution.

Ideally, however, our target should be documentary fieldwork. But how do we prepare young linguists to conduct such work? At UHM, our response is an endeavor that I will call the *Language Documentation and Conservation Initiative*. I will discuss this initiative in terms of (a) academic training, (b) collaborative research, and (c) open communication.

**3. ACADEMIC TRAINING.** In response to the need for linguists capable of conducting high-quality documentary fieldwork, UHM has established a graduate program in “language documentation and conservation” (LDC), one of the first of its kind in the United States.<sup>7</sup> Since its inception in 1963, this department has had a special focus on Pacific and Asian languages. It has supported and encouraged fieldwork in this region, and it has played a major role in the development of vernacular language education programs in Micronesia and elsewhere. The LDCI thus represents a renewed and intensified commitment to such work.

Our ideas about what we should be doing if our current focus is on endangered and/or underdocumented languages are still fluid, but it seems clear that linguists going into the field need training in at least five general areas. They need a solid foundation in (1) linguistic theory, (2) fieldwork methods and technology, (3) methods of language conservation, (4) area studies, and (5) what I will call, for lack of a better label, field skills, a category that includes knowledge of ethics, health, hygiene, and other capabilities that contribute to a fieldworker’s well-being. Our program is not equally prepared to provide formal training in all five of these areas—probably no program is—but at present our program is designed

---

<sup>7</sup> The URL for the LDC program is <http://www.ling.hawaii.edu/graduate/degreesandrequirements.html>.

as follows.

**3.1 COURSE STRUCTURE.** Our department offers a large number of graduate level courses in linguistics. The following chart thus lists just those that provide training in one or more of the five areas listed above. The check mark(s) listed after each course designate the *primary* goal(s) of the course.<sup>8</sup>

TABLE 1: Courses for language documentation and conservation students.

Course Number	Course Title	(1) Theory	(2) Fieldwork	(3) Conservation	(4) Areal	(5) Field Skills
410	Phonetics	•				
420	Morphology	•				
421	Phonology	•				
422	Grammar	•				
640G	Methods of Language Documentation		•	•		•
750G	Language Planning			•		
630	Field Methods		•			•
611	Acoustic Phonetics	•	•			
631	Language Data Processing		•			
640G	Polynesian Language Family				•	
640G	Language Contact	•				
640G	Anthropological Linguistics	•	•			
640S	Sociolinguistics	•	•			
645	Comparative Method	•				
661	Proto-Austronesian				•	
750F	Phonetic Fieldwork on Endangered Languages		•			
750G	Lexicography		•			
770	Areal Linguistics				•	

The first six courses listed above are required of all students in the LDC MA track. The seventh course, field methods or its equivalent, is required of all PhD students. The remaining courses are electives; MA students in the LDC track must take at least three of these. Note that the content of the 770 seminar in Areal Linguistics varies from semester

<sup>8</sup> Several courses are labeled 640G. This is a generic course number for 'topics in linguistics.'

to semester. Recent offerings have included courses on the Austronesian and Austroasiatic language families, as well as on the languages of Southeast Asia, the Pacific, the Central Pacific, Micronesia, and Borneo. Capable students who complete the MA program are encouraged to apply to the PhD program. At this level, the course requirements are flexible; hence, no special track is necessary.

Many of the courses listed in Table 1 are likely to be offered in any modern department of linguistics. What distinguishes the UHM program from many others are courses like *Methods of Language Documentation* (which provides an introduction to language documentation and conservation), *Language Planning* (not commonly offered in American universities), *Language Data Processing* (with a focus on corpus linguistics), *Phonetic Fieldwork on Endangered Languages* (of transparent importance) and *Lexicography* (which, so far as I am aware, is not regularly offered in any linguistics department in the United States). Linguistic theory also plays a central role in the UHM program, for reasons discussed next.

**3.2 THE ROLE OF LINGUISTIC THEORY.** It is an unfortunate fact that during the past four to five decades a schism has developed between linguists who are primarily theory-oriented and those who place a high value on fieldwork. This situation is perhaps nowhere more apparent than in our own field of Austronesian linguistics, where we now have two competing conferences, both of which originated in the United States—the ICAL (International Conference on Austronesian Linguistics) series, which are broad in scope and attract many fieldworkers, and the AFLA (Austronesian Formal Linguistics Association) series, which focus on the contributions that Austronesian languages might make to linguistic theory.<sup>9</sup> While there are, of course, some linguists who attend both conferences, the question remains, why does this schism exist? An answer to this question is well beyond the scope of this paper, but one observation about fieldwork is worth noting.

In many departments of linguistics, perhaps especially in the United States, fieldwork and fieldworkers have unfortunately, but unquestionably, been marginalized. In 2004, Paul Newman surveyed 45 American universities offering a PhD in linguistics and found that, while 80% offered a course in field methods, only 38% required it of PhD students and only 42% offered the course every year.<sup>10</sup> These statistics were virtually unchanged from a similar survey he had carried out in 1992. The unfortunate fact is, in many American universities fieldwork is not encouraged, and, if it *is* undertaken, it is often for the purpose of gathering limited data germane to a specific theory; it is what I have called artificial fieldwork.

Clearly, this situation is deplorable, but there is no reason why it need continue. In fact, there is at present an increasing concern for healing this breach and for strengthening the empirical foundation of our discipline, triggered largely by concerns about language endangerment and loss. The 2007 winter meeting of the Linguistic Society of America, for example, included a symposium on “Endangered Languages and Linguistic Theory”,

<sup>9</sup> There are, of course, other regularly scheduled conferences that focus on Austronesian languages, including the Austronesian Languages and Linguistics (ALL) series, and the Conference on Oceanic Languages (COOL) series.

<sup>10</sup> See Newman 2005.

as well as an address by Mark Liberman on “The Future of Linguistics,”<sup>11</sup> in which he suggested new directions for our discipline, including a focus on language description and documentation.

At UHM, our goal is bring data, documentation, analysis, description, and theory together in one seamless whole. We do not diminish the importance of any of these elements. We believe such an approach to be good science; the creation of a useful record of a language necessarily entails a considerable amount of analysis<sup>12</sup>, and linguistic analysis requires linguistic theory. In our case, however, providing students with a solid grounding in linguistic theory is also a necessity. As Peter Austin (2003:10) has observed: “it is important that we think about how language documenters can advance their careers, or at least not set them back.”<sup>13</sup> We work within an American academic setting, and if our intent is to produce students who are employable—and that should be a primary concern of every department—it is essential that our graduates be able to communicate in the language of mainstream American linguistics.<sup>14</sup>

Linguistic theory is thus an essential component of our program, but we attempt to ensure that our students are not theory-bound. We encourage exposure to competing theories, and we try to ensure that our students understand that there are more wonders among the languages of this planet than are dreamed of in the halls of academia. In fact, it is not uncommon, I suspect, for fieldworkers to experience what I call theory-lag. That is, they not infrequently observe phenomena for which no current theory has any satisfying explanation.<sup>15</sup> In this respect, it is good fieldwork that can put one on the cutting edge of linguistic theory.

At UHM, we fully recognize that the theories, and for that matter the technology we work with, are ephemeral, but they cannot for that reason be ignored. As Thorstein Veblen (2004:197) has astutely observed: “Invention is the mother of necessity.”<sup>16</sup>

**4 COLLABORATIVE RESEARCH.** The field of language documentation is evolving into a borderless discipline. It is therefore essential that linguists forge alliances with all those who have a stake in the documentation and conservation of linguistic, cultural, and biological diversity. The question is, how do we involve others in this essential work?

Clearly, the people who have most at stake in the documentation and maintenance

<sup>11</sup> The slide show for this talk can be found at <http://ldc.upenn.edu/myl/LSATalk.pdf>.

<sup>12</sup> For example, a linguist cannot competently transcribe and annotate a text unless s/he knows a considerable amount about the phonology, morphology, and syntax of the language.

<sup>13</sup> At a recent conference, I was told by a faculty member from an American university that she would strongly discourage young linguists from getting involved in language documentation until they had tenure. Clearly, until we can fully professionalize our work, we must find ways to overcome such concerns.

<sup>14</sup> It is also important that linguists, both students and professionals, be able to recognize data of theoretical/ typological interest when they encounter it.

<sup>15</sup> For examples, see Blevins 2007.

<sup>16</sup> The full quote from Veblen is: “And here and now, as always and everywhere, invention is the mother of necessity”.

of threatened languages are the people who speak them. Part of the training of young linguists, then, ought to include experience doing collaborative research with speakers of such languages. At UHM, the key effort in this regard has been the *Language Documentation Training Center*, described next.

**4.1 LANGUAGE DOCUMENTATION TRAINING CENTER.** The *Language Documentation Training Center* (LDTC)<sup>17</sup> was initiated by Meylysa Tseng, a linguistics graduate student at UHM. In 2003, Ms. Tseng set out to organize a community service project. Given the linguistics department's long-standing interest in documenting languages, and given the rich cultural and linguistic diversity present on the campus, she decided upon an activity that would bring together graduate students in linguistics and speakers of minority and underdocumented languages. The result was the project now known as the LDTC.

The mission of the LDTC, as described by the students who run it, is to equip native speakers with rudimentary skills in documentation, to offer them a public outlet for information about their languages, to support them in their language analysis and documentation efforts, and to inspire them to become language advocates in their own communities.

To accomplish these goals, international students partner with graduate students in linguistics in a one-semester training program that consists of eight two-hour workshops conducted on Saturday mornings.<sup>18</sup> These workshops, led by volunteer graduate students, familiarize the participants with endangered language issues and provide training in basic documentation skills, including digital recording, lexicography, and translation. The final product at the end of the semester is a webpage that includes basic information about the student's language. At present, these webpages are structured to include (1) biographical information about the student and basic information about the language (where it is spoken, number of speakers, etc.), (2) rudimentary information about the sound system, (3) a rendition of the Bird Story<sup>19</sup> in the target language, with morpheme-by-morpheme glosses, a free translation, and a sound file, (4) brief comments on the morphology and syntax of the language, (5) a two-hundred-word Swadesh list, including sound files for at least 25 of the items, plus (6) information on the orthography.<sup>20</sup> To date, the project has produced 44 webpages. Because the project is open to all students, some of the pages are devoted to relatively well documented languages that are not endangered, such as Javanese, but the LDTC site nevertheless includes a remarkable range of languages, including some for which there is little or no other documentation, thus highlighting the rich linguistic environment in which our students work.

It is difficult to overstate the impact that the LDTC has had on our department. It has provided a laboratory for our documentation students, it has resulted in a stirring sense of camaraderie among its participants, and it has been well-received, both on the campus and

---

<sup>17</sup> The URL for the LDTC is <http://www.ling.hawaii.edu/~uhdoc/index.html>.

<sup>18</sup> The partners in the LDTC often work together outside of workshop hours as well. Participants are encouraged to return in following semesters to delve into their languages more deeply.

<sup>19</sup> Based on a wordless picture book. See the LDTC site for details.

<sup>20</sup> See, for example, the website for Manadonese at <http://www.ling.hawaii.edu/~uhdoc/manadonese/>.

in the community. It is a project with goals that others can understand, perhaps especially in Hawai'i, where the indigenous language is highly endangered. The LDTC has also been given a substantial amount of publicity in newsletters, newspapers, and on the radio. In 2005, the center received the national "NAFSA-TOEFL-ETA Partnership in Excellence Award" for its innovative methods of involving international students in on-campus activities. In that same year, it also won first place in the UH College of Business "Business Plan Competition" in the non-profit "social" category, reflecting the students' goal of making the project self-supporting. More recently, in Fall 2006, the LDTC won first place in the UH Sustainability Awards in the category of "cultural conservation."<sup>21</sup>

The activities of the LDTC have also led to a number of important spin-offs. For example, several of our current students have been working with a community of Tokelauans who live on O'ahu. They are assisting in the development of a Tokelauan learner's dictionary for elementary school students, as well as a corpus of written and spoken Tokelauan. Two of our faculty members have also assisted the Tokelauans in designing and administering a questionnaire to learn more about the use of Tokelauan in the local community. Most recently, the LDTC has been invited to work with speakers of minority languages at one of the largest high schools in Honolulu. This activity will open up new opportunities for the center and allow our graduate students to become more directly involved with the community.

The accomplishments of the LDTC have also served as a catalyst for involving other departments and organizations on the campus in the important work of language documentation and conservation. In 2006, for example, four federally funded centers on our campus—the National Foreign Language Resource Center, the National Resource Center for East Asia, the Center for Pacific Island Studies, and the Center for Southeast Asian Studies—wrote proposals for renewed funding. All of these centers requested and received funds to support the UHM language documentation and conservation initiative. The LDTC, then, has been a very successful program with far-reaching consequences. Where feasible, I would urge other linguistics departments to consider establishing a comparable student-directed center. The payoffs can be substantial.

**4.2 FUTURE COLLABORATIVE ACTIVITIES.** At present, UHM is planning three additional activities that focus primarily on collaborative research and learning. These are (1) a conference, (2) a summer institute, and (3) the creation of an alliance of people concerned with issues of language and cultural sustainability.

In 2009, UHM will host a conference on language documentation and conservation. While we have not yet decided upon a theme for this conference, collaborative research is likely to be a major focus. The loss of a language affects not only its speakers and linguists, but anthropologists, archeologists, ethnobotanists, ethnomusicologists, folklorists, historians, ichthyologists, ornithologists and many others as well. Linguists have assumed the burden of documenting endangered languages, but it is essential that we involve others whose lives and careers are affected by the current world-wide convergence of language and culture.

In 2010, UHM will host a summer institute designed to bring together language ac-

---

<sup>21</sup> See <http://sustainable-uh.hawaii.edu/index.php?section=2&island=2>.

tivists, graduate students in linguistics, professional linguists, and others with interests in language documentation and/or conservation. The planning for this institute is still in the preliminary stages, but it is likely that it will provide instruction in field methods and related topics, including orthography development, dictionary and reference grammar design, and audio and video recording techniques. It will additionally include training in language maintenance and revitalization work, including the development of literacy materials and the utilization of current technologies to gather, process, disseminate, and archive linguistic information. This institute will be the second in what we hope will become a series of such institutes, to be held every two years. The first such institute, named InField (Institute on Field Linguistics and Language Documentation), will be held in 2008 at the University of California in Santa Barbara.<sup>22</sup> Present plans call for the University of Oregon to host the institute in 2012.

In conjunction with the conference and summer institute, faculty and students at UHM are also exploring how they might create a local alliance of people interested in the work of language documentation and conservation. The current climate for such an effort on the Mānoa campus is excellent. For example, the university has established an "Office of Sustainability," which posts as its theme "A sustainable university, reflecting traditional island values."<sup>23</sup> In the first UHM Sustainability competition, 29 projects were submitted from across the campus, representing a wide variety of interests and disciplines. As previously noted, the student-directed LDTC project won one of the eight awards.

Clearly, there is much to be gained by cooperating with others. We view such efforts as essential to our program, in large part as a consequence of what we perceive as our need to overcome problems of insularity—the insularity of our location, of our discipline, and of academia in general. We believe that these efforts will not only enhance the stature of our program, but of our discipline as well. In this respect, the language documentation and conservation initiative at UHM is establishing a new agenda in which our students are urged to become involved with others in shaping our local, regional, and global future. If we are going to attempt to meet the challenge of documenting and sustaining hundreds, or even thousands, of languages over the course of this century, then such collaborative efforts are essential. Time is short, and linguists cannot do it all.

**5. OPEN COMMUNICATION.** Paul Newman (2007:28) has noted that an alternative reading for the acronym LD&C (Language Documentation and Conservation) might be Language Documentation and Communication, a reminder to us all that disseminating the results of our research is as essential as the research itself. All too often, the insights we glean and the information we gather remain buried in our offices.<sup>24</sup> Thus, two of the goals of the language documentation and conservation initiative at UHM are (1) to provide a publishing outlet for field linguists, and (2) to establish an archive where linguistic data and documents can be safely stored and, consistent with the wishes of the speech community, made available to others.

---

<sup>22</sup> See <http://www.linguistics.ucsb.edu/faculty/infield/>

<sup>23</sup> See <http://sustainable-uh.hawaii.edu/index.php?section=1>

<sup>24</sup> I do not exclude myself from this accusation.



**5.1 THE LANGUAGE DOCUMENTATION AND CONSERVATION JOURNAL.** Not publishing the results of one’s fieldwork is a serious problem. However, publishing our work in exorbitantly priced books and journals seems to me a scant improvement upon having not published them at all.

While there are many publishing outlets for research that is theory-oriented, or even descriptive-oriented, until now there has been no journal devoted to the wide range of interests of field linguists. UHM is attempting to fill this need by launching *Language Documentation and Conservation (LD&C)*, a new, fully-refereed, open-access journal that is sponsored by the National Foreign Language Resource Center and published exclusively in electronic form by the University of Hawai‘i Press.<sup>25</sup>

The homepage for this journal notes that: “*LD&C* publishes papers on all topics related to language documentation and conservation, including, but not limited to, the goals of language documentation, data management, fieldwork methods, ethical issues, orthography design, reference grammar design, lexicography, methods of assessing ethnolinguistic vitality, archiving matters, language planning, areal survey reports, short field reports on endangered or underdocumented languages, reports on language maintenance, preservation, and revitalization efforts, plus software, hardware, and book reviews.”

*LD&C* is designed to ensure that it is available to the widest audience possible. Therefore, the journal is free and open to all, without the need to subscribe. The choice of an electronic format for this journal allows it to include audio and video content, links to other sites, and, because its content is available in HTML, as well as PDF, it is also accessible to the handicapped.

**5.2 FUTURE ACTIVITIES.** If the products of our documentation efforts are not properly archived and made accessible to others, now and in the future, our efforts will have been for naught. Consequently, it is our intention to establish a language archive center at UHM, to be directed by Nicholas Thieberger, who will be joining our faculty in January of 2008.

We also wish to explore ways in which we can improve upon the design of the products we create and store. That is, while we are very much concerned with the processes of language documentation, we also intend to pay increasing attention to what we produce.

If our work is to be of use to others, it is also essential that we give thought to how we might create linguistic products that are low-cost or free, accessible and comprehensible to the widest possible audience, sensitive to the needs of the speakers of the target languages, and of maximal value in the future. Speakers of threatened languages commonly want linguists to assist them in developing basic literacy tools—orthographies, dictionaries, reference grammars, and reading materials. But what form should these products take? What constitutes an optimal orthography, and, given the rapid changes taking place in the media, what form will dictionaries, grammars, and reading materials take in the future? These are important concerns that will require additional consideration from those of us interested in assisting minority language communities.

---

<sup>25</sup> The URL for this journal is <http://nflrc.hawaii.edu/ldc/>. The inaugural issue appeared in June, 2007.

**6. CONCLUSION.** Let me end this paper on a personal note. As a result of our endeavors in the areas of language documentation and conservation, our department has been obligated to come to terms with what I see as three common shortcomings of linguistics programs in the United States. These are (1) the failure to train young linguists, perhaps most especially international students, to gather original data, (2) the failure to involve others in our research efforts, and (3) the failure to consider how we might make at least some of our research findings useful and accessible to a broad audience.

The greatest benefit of the language documentation and conservation initiative to our department, however, has been its effectiveness as a recruiting tool. It has attracted excellent students who are active, engaged, enthusiastic, and committed to what they are doing. In many respects, they remind me of the young people I worked with in the Peace Corps in the 1960s. These students desire to live purposefully, and they are unafraid to accept challenges. They have revitalized our department.

A second benefit to our department has been the recognition we have received for our efforts, not so much from those within our discipline, but from those outside it. Our activities in the community, as well as the many students we have sent into the field, have helped to create an image of a department that is doing important and consequential work, and that is playing a meaningful role in society.

Finally, we work with the hope that we are heeding Jonas Salk's admonition—that we try to learn to be good ancestors. Some linguists have downplayed the significance of language loss, arguing that because languages change, new languages will evolve. But this strikes me as specious reasoning. Speaking of the loss of biological diversity, E. O. Wilson (2006:84) has observed that, after a mass extinction of species: "The original level of biodiversity is not likely to be regained in any period of time that has any meaning for the human mind." If we lose more than half of the world's languages this century, how long will it take to regain the current level of linguistic diversity? 10,000 years? 50,000 years? Ever? Wilson (2006:55) also teaches us that, "If a miracle is a phenomenon we cannot understand, then all species are something of a miracle". And so, too, are all languages.

## REFERENCES

- AUSTIN, PETER K. 2003. Introduction. In *Language documentation and description*. Vol. 1, ed. by Peter K. Austin, 6–14. (Endangered Languages Project). London: School of Oriental and African Studies.
- BLEVINS, JULIETTE. 2007. Endangered sound patterns: Three perspectives on theory and description. *Language Documentation & Conservation* 1(1): 1-16. Honolulu: University of Hawai'i Press. <http://nflrc.hawaii.edu/ldc/>.
- HIMMELMANN, NIKOLAUS P. 2006. Language documentation: What is it and what is it good for? In *Essentials of language documentation*, ed. by Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, 1–30. Berlin: Mouton de Gruyter.
- KRAUSS, MICHAEL E. 1992. The world's languages in crises. *Language* 68(1): 4–10.
- NEWMAN, PAUL. 1992. Fieldwork and field methods in linguistics. *California Linguistic Notes* 23(2): 2–8.
- NEWMAN, PAUL. 2005. Field methods courses in linguistics. Paper presented at the Linguistic Society of America conference on Language Documentation: Theory, Practice, and Values. July 9–11, Harvard University.
- NEWMAN, PAUL. 2007. Copyright essentials for linguists. *Language Documentation & Conservation* 1(1): 28-43. Honolulu: University of Hawai'i Press. <http://nflrc.hawaii.edu/ldc/>.
- RHODES, RICHARD A., LENORE A. GRENOBLE, ANNA BERGE, and PAULA RADETZKY. n.d. *Adequacy of documentation: A preliminary report to the CELP* [Committee on Endangered Languages and their Preservation, Linguistic Society of America].
- VEBLEN, THORSTEIN. 2004. *The instinct of workmanship and the state of the industrial arts*. Whitefish, MT: Kessinger Publishing. [First published in 1914.]
- WILSON, E. O. 2006. *The creation: An appeal to save life on Earth*. New York: W.W. Norton & Company.

Kenneth L. Rehg  
rehg@hawaii.edu

# 3

## **Training for Language Documentation: Experiences at the School of Oriental and African Studies**

Peter K. Austin

*School of Oriental and African Studies, University of London*

Since 2003 the Endangered Languages Project at SOAS has been involved in various types of training for documentation of endangered languages, ranging from one-day workshops through to MA and PhD post-graduate degree programs. The training events have been attended by specialists, research grantees, students, and members of the general public, and have covered a wide range of topics and involved delivery in a range of contexts and delivery modes, including hands-on practical sessions and e-learning in the Blackboard framework. We have covered both theory and practice of language documentation and endangered language support, including the development of multimedia and curriculum materials for language teaching, some of it experimental and, we think, quite innovative. In this paper I discuss some of our experiences in developing and running these training workshops and courses, reporting on the models, and successes (and failures) over the past three and a half years. My goal is to share our accumulated knowledge and experience with others with similar interests, and in doing so to advance our understanding of the possibilities for language documentation training.

**1. INTRODUCTION.**<sup>1</sup> The last ten years has seen the emergence of the new field of Documentary Linguistics as an area of research that deals with the principles and practices of documenting languages, particularly endangered languages. Correspondingly, a need has developed for training both practicing linguists and a new generation of students in the theory and methods of language documentation. This paper is an overview of the training programs we have set up at the School of Oriental and African Studies (SOAS), University of London, to address this need. Section 2 outlines documentary linguistics and language documentation as generally conceived, and Section 3 identifies the range of skills required to undertake language documentation. Section 4 introduces the Hans Rausing Endangered Languages Project (HRELP), and in Section 5 I discuss the training models we have developed at SOAS to introduce and extend language documentation knowledge and skills for post-graduate students and established linguists who have received a grant from HRELP. In Section 6 I give an assessment of some of the successes and challenges we have seen over

---

<sup>1</sup> This paper was written for the International Conference on Austronesian Endangered Language Documentation, Providence University, Taiwan 5-7 June 2007. The ideas and materials presented here have been discussed in some detail with Oliver Bond, Lenore Grenoble, Colette Grinevald, Anthony Jukes, Friederike Luepke, Sophie Manus, Robert Munro, and David Nathan. I am grateful for feedback on an earlier version from two anonymous referees. None of these people can be held responsible for any errors or shortcomings remaining.

the five years HRELP has been operating. Section 7 presents some conclusions from our experience that may be of interest and value to others wishing to establish training courses for language documentation.

**2. DOCUMENTARY LINGUISTICS AND LANGUAGE DOCUMENTATION.** Documentary linguistics is a newly emerging field of linguistics that is “concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties” (Gippert, Himmelmann and Mosel 2006:v). Documentary linguistics has developed over the last decade in large part in response to the urgent need to make an enduring record of the world’s many endangered languages and to support speakers of these languages in their desire to maintain them (Whalen 2003, Austin 2007). It is also fueled by developments in information, communication and media technologies which make documentation and the preservation and dissemination of language data possible in ways which could not previously be envisioned. In addition it essentially also concerns itself with the roles of language speakers in documentary projects and their rights and needs in ways not previously considered within linguistics (see Thieberger and Musgrave 2007).

Himmelmann (2006:15) identifies several important new features of documentary linguistics:

- Focus on primary data – language documentation concerns the collection and analysis of an array of primary language data to be made available for a wide range of users;
- Explicit concern for accountability – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected;
- Concern for long-term storage and preservation of primary data – language documentation includes a focus on archiving in order to ensure that documentary materials are made available to potential users into the distant future;
- Work in interdisciplinary teams – documentation requires input and expertise from a range of disciplines and is not restricted to linguistics alone;
- Close cooperation with and direct involvement of the speech community – language documentation requires active and collaborative work with community members both as producers of language materials and as co-researchers.

I use the term language documentation to refer to the activities carried out by researchers and communities engaged in work that adopts a documentary linguistic approach. The historical genesis of the field of documentary linguistics has meant that the term ‘language documentation’ is sometimes used loosely, to refer to any kind of language record, but documentary linguistics uses it in a more specific way, to refer to an activity with much larger and more specific goals. In particular, language documentation strives “to provide a comprehensive record of the linguistic practices characteristic of a given speech community” (Himmelmann 1998:166). Language documentation differs fundamentally and critically

from language description. Language documentation seeks to record the linguistic practices and traditions of a speech community, along with speakers' metalinguistic knowledge of those practices and traditions. This includes systematic recording, transcription, translation and analysis of the broadest possible variety of spoken (and written) language samples collected within their appropriate social and cultural context (Austin 2006, HRELP 2006). Analysis within language documentation is aimed at making the records, or rather the language data recorded, accessible to a broad range of potential users. This group includes not only linguists but also community members, who may not have first-hand knowledge of the documented language. The record is thus intended for posterity, and so some level of analysis is required, in particular glossing and translation into one or more languages of wider communication (see Evans and Sasse 2007 for some of the challenges that entails), and systematic recording of metadata to make the archived document(s) findable and usable (Nathan and Austin 2004).

I take the core of a language documentation project to be the creation of a corpus of audio and/or video materials with time-aligned transcription, multi-tier annotation, translation into a language of wider communication, and relevant metadata on context and use of the materials. Woodbury (2003) argues that the corpus will ideally be large, cover a diverse range of genres and contexts, be expandable, opportunistic, portable, transparent, ethical and preservable. As a result, documentation is increasingly done by teams, including community members, rather than 'lone wolf linguists;' both the technical skills and the amount of time required to create this corpus make it difficult for a single linguist, working alone in the field, to achieve.

Language documentation typically begins with the development of a project to work with a speech community on a language and can be seen as progressing through a series of stages, some of which are carried out in parallel:

- Project conceptualization and design
- Establishment of field site, including negotiation of permissions
- Funding application
- Data collection and processing
- Creation of outputs
- Evaluation and reporting

The following stages in the data collection and processing phase can be recognized (Austin 2006):

1. *recording* – of media (audio, video, image) and text
2. *capture* – moving analogue materials to the digital domain
3. *analysis* – transcription, translation, annotation, and notation of metadata

4. *archiving* – creating archival objects, and assigning access and usage rights

5. *mobilization* – publication, and distribution of the materials in various forms

Language description typically involves the production of grammars, dictionaries, and collections of texts<sup>2</sup>. In contrast, the primary goal of language documentation is the development of a corpus which is representative of a wide range of discourse types (Austin 2006, Woodbury 2003, Himmelmann 1998). Although description relies on documentation (and documentation essentially includes descriptive aspects such as annotation), it involves analysis of a different order: description provides an understanding of language at a more abstract level, as a system of elements, rules, constructions and so on (see again Himmelmann 1998, 2002:48). Description and analysis are contingent by-products of documentation and will change and develop over time as research progresses (Woodbury 2003, Austin and Grenoble 2007). Such works can be valued by speech communities and provide important input into processes of language maintenance and revitalization, however, the primary audience for these products is typically linguists, and sometimes they are written in frameworks accessible only to trained linguists. Such products can also become unusable as linguistic theoretical models come into and go out of fashion. Thus many of the grammars written in the 1970's and 1980's in Tagmemic or Transformational Grammar frameworks are now extremely difficult to use. Language documentation focuses specifically on providing a preservable and transparent corpus of analysed materials on a language that is well structured and designed for access by non-linguists as well.

**3. SKILLS FOR LANGUAGE DOCUMENTATION.** Language documentation requires knowledge and application of a range of skills, including those traditionally associated with fieldwork and language description, as well as skills in the application of information, communications and media technologies (Munro 2005) and applied ethics. Increasingly also, documenters are expected to have knowledge and skills typically associated with areas of applied linguistics, such as orthography development, lexicography, translation, pedagogy and curriculum design, multimedia, language policy and needs assessment, and advocacy. The need for these skills arises from the desires and expectations of the language communities and the multidisciplinary orientations of the work.

I propose the following is an (incomplete) list of documenter skills that researchers should have some exposure to and competence in:

- *Project conception, design and management* – familiarity with documentation theory, applied ethics, intellectual property rights and socio-cultural issues, stakeholder communication
- *Grant application writing*

---

<sup>2</sup> Historically, and in some cases currently, some linguists use the term 'language documentation' to refer to what we are calling 'language description'. I attempt to be consistent in my usage of the two terms.

- *Media management* – recording techniques, field methods, data transfer, backup
- *Data and metadata management* – data and metadata representation (XML, relational database models), transcription, linguistic analysis (phonetics, phonology, morphology, syntax, semantics) and annotation, use of linguistic software tools (Transcriber, Shoebox/Toolbox, ELAN, IMDI), data integrity and sustainability (Bird and Simons 2003), workflow design and management
- *Mobilization* – familiarity with applied linguistics concepts (orthography design, lexicography, curriculum development, policy formation, revitalization), publication skills, multimedia design and implementation (Csato and Nathan 2003, Nathan 2006, Nathan and Csato 2006)
- *Team-based research* – skills sharing and transfer, capacity development
- *Reporting* – presentation, writing and communication skills

It is highly unlikely that any single researcher will have all this knowledge and skills at a sufficient level to be able to apply them all in carrying out a project. Indeed, as Walcott (1999:73) says of ethnography:

“instead of envisioning the ethnographer as Superman or Wonder Woman, one must recognize that it is the scope of the ethnographic question that must be pared to what one individual, or a researcher working with a colleague or small research team, can accomplish in a limited amount of time. The range of fieldwork techniques that can be employed must be pared as well. Of course, in an ideal world, every researcher would be sufficiently talented to be able to summon from a vast personal repertoire whatever combination of techniques seems appropriate to addressing the issue at hand.”

In other words, projects need to be realistic in terms of what can be achieved in the time available, and researchers must ensure that they acquire the skills they need for that particular project. Indeed, Walcott (ibid) argues that:

“if you, as sole or principal investigator, really did have all the skills of social research at your command – computer skills, language skills, statistical skills, survey techniques, ability to work with experimental and quasi-experimental design – let alone all the observer and interviewer and interpersonal skills an ethnographer is likely to need, why would you invest your time plodding along with ethnography?”

We just need to substitute ‘language documentation’ for ‘ethnography’ here to make his advice applicable to our current concerns.



**4. THE HANS RAUSING ENDANGERED LANGUAGES PROJECT.** The Hans Raising Endangered Languages Project (HRELP) was established with a commitment of £20 million from Arcadia (formerly the Lisbet Raising Charitable Fund) to document as many endangered languages as possible and to encourage the development of relevant skills across the world. It has the following three components:

1. *Endangered Languages Documentation Program (ELDP)* – will provide approximately £15 million over an 8-10 year period in competitive research grants to encourage the development of linguistic fieldwork in endangered languages (especially by younger scholars) and to support documentation of as many threatened languages as possible. ELDP offers five types of grants, and is governed by an international selection panel chaired by Prof. Graham Furniss of the School of Oriental and African Studies (SOAS); the ELDP grants administration is managed by SOAS.

2. *Endangered Languages Academic Program (ELAP)* – aims at training the next generation of language documenters, it offers students an array of opportunities: an MA in Language Documentation and Description, a PhD in Field Linguistics, and post-doctoral fellowships at SOAS. We also offer a comprehensive program of public lectures, seminars, workshops and training courses. Prof. Peter K. Austin, Märit Raising Chair in Field Linguistics, is Director of ELAP.

3. *Endangered Languages Archive (ELAR)* – is building up a large collection of endangered languages documentation resources, and also supports training, development of methodologies, and other technical aspects of language documentation and archiving. David Nathan is Director of ELAR.

HRELP has involvement in 18 research projects concerned with endangered Austronesian languages, as set out in the following Table<sup>3</sup>. All projects except one are funded by ELDP and some involve post-graduate students or post-doctoral researchers who are being trained by ELAP; all other grantees have attended training courses at SOAS (see 5.2 below). All the data collected by these projects is being archived at ELAR.

TABLE 1. HRELP: Endangered Austronesian Languages Research Projects

Location	Language	Researcher & Institution <sup>1</sup>	Notes
Taiwan	Yami	Der-Hwa Victoria Rau, Providence University, Taiwan	grant 2005-07
Philippines	Palawan	Charles Macdonald, CNRS, France	grant 2006-07
Micronesia	Pingilapese	Ryoko Hattori, University of Hawai'i at Mānoa, USA	grant 2006-07

<sup>3</sup> For further details see <http://www.hrelp.org/grants/projects/>

Indonesia	Toratán	Anthony Jukes, SOAS	grant 2005-07, ELAP post-doc
	Helong	John Bowden, ANU, Australia	grant 2007-10
	Rongga	Wayan Arka, ANU, Australia	grant 2004-06
	Allang Amahai Tulehu	Margaret Florey, Monash University, Australia	grant 2003-05
East Timor	Maku'a	Aone van Engelenhoven, Leiden University, Netherlands	grant 2003
Malaysia	Eastern Penan	Peter Sercombe, Northumbria University, UK	grant 2003-04
New Caledonia	Numèè	Sophie Rendina, SOAS	grant 2007-09, ELAP PhD student
Vanuatu	Mafea	Valérie Guérin, University of Hawai'i at Mānoa, USA	grant 2005-07
	Vurès	Catriona Malau (nee Hyslop), La Trobe University, Australia	grant 2004-06
	Biriebo	Peter Budd, SOAS	grant 2005-07, ELAP PhD student
	Seke	Kay Johnson, SOAS	grant 2007-09, ELAP PhD student
	Neverver	Julie Barbour, Waikato University, New Zealand	grant 2004-07
	Qatareu	Hans Schmidt, Hamburg University, Germany	grant 2004-06
Solomon Islands	Kubakota	Mary Raymond, SOAS	ELAP PhD student
	Blablanga	Radu Voica, SOAS	grant 2007-09, ELAP PhD student

**5. TRAINING FOR LANGUAGE DOCUMENTATION AT SOAS.** At SOAS we have developed several types of training for scholars wishing to undertake language documentation. This training is run by ELAP in collaboration with ELAR – some is directly funded by ELDP:

1. post-graduate courses: MA (1 year), PhD (3-4 years)

2. grantee training courses: one week
3. specialized training in recording, archiving, XML: 1 day

**5.1 POST-GRADUATE COURSES.** We offer two levels of post-graduate courses: a one year Master of Arts degree, and a three-year PhD. Students entering the PhD must have already completed an MA or equivalent.

The MA in Language Documentation and Description is a one year degree that is taught over 20 teaching weeks. The intake includes students who hold an undergraduate BA degree with a major in linguistics, along with those with no previous exposure to linguistics (generally in a proportion of 1/3 with a linguistics major and 2/3 without). The degree consists of course-work plus a short dissertation (10,000 words, essentially a long research essay), and includes a Research Training Seminar (1 hour throughout the year) and recommended attendance at fortnightly departmental seminars and occasional workshops and training courses that are held throughout the year. Starting in 2007-08 there are two pathways in the degree: a Field Linguistics pathway for students with an undergraduate major in linguistics or equivalent, and a Language Documentation and Support pathway for students with or without linguistics. The following tables summarize the structure of the two pathways.

TABLE 2. Field Linguistics Pathway

	Core	Core	Core/Option	Core
Term 1	Issues in language documentation	Technology & language documentation	Option 1	Research Foundations
Term 2	Field methods	Applied documentation & description	Option 2	Research Foundations

TABLE 3. Language Documentation and Support Pathway

	Core	Core	Core/Option	Core
Term 1	Principles of linguistic analysis	Issues in language documentation	Option 1	Research Foundations
Term 2	Principles of linguistic analysis	Applied documentation & description	Option 2	Research Foundations

The core courses are the following:

- *Principles of linguistic analysis* – an introduction to basic linguistic concepts and analytical techniques with hands-on practical exercises

- *Issues in language documentation* – this covers the topics of: project design, research ethics, IPR, researchers and communities, world language ecology, endangered and minority languages, language contact, ethnographic methods
- *Technology and language documentation* – covers information, communications and media technology basics, audio and video recording and editing techniques, data analysis and design, data formats and standards, archiving issues, software tools (Transcriber, Shoebox ~ Toolbox, ELAN, Praat, IMDI) and practical exercises
- *Applied language documentation* – introduces basic principles of applied linguistics, including orthography design, literacy, lexicography, translation, language learning and teaching, pedagogy, curriculum design, language policy, and advocacy
- *Field methods* – students work with a speaker of an unknown language to apply their knowledge and skills of data recording and analysis. In recent years the languages studied have been Khorchin Mongolian, Dida, Sylheti, and Kannada.

The Research training seminar consists of weekly meetings attended by all the students together to explore conceptual issues (e.g. what counts as explanation in linguistics, types of data, brief history of linguistic research), research methods and skills, research tools (End-Note, Powerpoint), presentation skills (including presentation dry runs), and teamwork skills. This class is especially important for cohort development and providing a forum for student concerns (e.g. workload, assessment) to be aired and addressed.

The options courses include linguistic typology, syntax, phonology, semantics, historical linguistics, language culture and society, acoustic and experimental phonetics, multimedia and language support, and areal courses (Austronesian, African languages, American Indian languages, Siberian languages). We plan to introduce a course on language revitalization in the near future as this is a topic students are keenly interested in. Students can also choose from advanced options primarily intended for PhD training (see below). These vary from year-to-year but have included grammar writing, negation, tense/aspect/mood, number, and lexical semantics.

The PhD generally takes 3 to 4 years and follows from the MA degree. SOAS offers two PhDs: a general linguistics PhD and specialist PhD in Field Linguistics. The following is the structure of this latter degree:

Year 1 – advanced level training in language documentation and description: students take 3 courses per term on advanced subjects, along with a special 3 day training course. They prepare a “core chapter” of the dissertation and an annotated bibliography, give a public seminar and are upgraded from MPhil to PhD if their work is of the required standard

Year 2 – fieldwork, normally 8-12 months

Year 3/4 – writing up dissertation, usually with a short fieldtrip (3-4 months) to check data and fill gaps

Examination of the PhD is based entirely on the dissertation and the coursework components of the degree are required but not assessed.

Occasionally, we have run special additional training events such as a three day ELAP/ELAR training program designed for PhD students and post-doctoral fellows which aim to:

- address language documentation workflow and supporting principles and skills
- develop a shared involvement in collaborative project work to enhance knowledge and skills of audio recording, processing and presentation

The following table shows the structure of the most recent such training course.

TABLE 4. ELAP/ELAR training program for PhD students and post-doctoral fellows

	Wednesday	Thursday	Friday
9.45-11am	“Managing My Data”	Audio 3: Evaluation	Images
11-11.15am	Tea break	Tea break	Tea break
11.15am-1pm	Working in teams	Audio 4: Editing	Archiving
1-2pm	Lunch	Lunch	Lunch
2-3.30pm	Audio 1	More data management	ELAN Transformation
3.30-3.45pm	Tea break	Tea break	Tea break
3.45-5.15pm	Audio 2: Practical	ELAN	Mobilization/Multimedia

**5.2 ELDP GRANTEE TRAINING. ELAP and ELAR,** in collaboration with outside specialists, offer a six-day training course for researchers who are awarded a grant by ELDP. The training course is designed to assist grantees by building on their knowledge of documentation theory and practices, especially information and media technology skills, and preparation for dealing with grant outcomes (corpus development, archiving, mobilization and publication). The following table sets out the structure of our training course run in June 2007 (the structure builds on two years of experience running these courses):

TABLE 5. ELDP grantee training program

	Wednesday	Thursday	Friday	Saturday	Monday	Tuesday
09:00 - 09:50	Welcome	Audio evaluation	Data management	Documentation	Video and documentation	Projects
9:50 - 10:40	Grantee projects				Videography	
10:40 - 11:00	Tea/Coffee					
11:00 - 11:50	Audio principles	ELDP relations	Data practical	ELAN	Video: camera	Mobilization
11:50 - 12:40		Archiving			Video: practical	
12:40 - 14:00	Lunch					
14:00 - 14:50	Digital audio	Archiving	Consultation & elicitation	Advice clinic	Video: practical	Ethics and IP
14:50 - 15:40	Audio practical	Transcription issues				
15:40 - 16:00	Tea/coffee					
16:00 - 16:50	Audio practical	Transcription practical	Field practical topics	Projects and questions	Video editing & evaluation	Wrap-up
16:50 - 17:40	Grantee projects					

Note: the shaded areas show sessions focusing on media recording, editing and management

We have run four of these training courses so far and all have had different dynamics (reflecting the experience and interest of the participants) but have all been evaluated by their participants as successful. Researchers generally feel that their existing knowledge and skills have been recognized and that they have acquired new information that they will put into practice in their own projects.

**5.3 SPECIALIZED WORKSHOPS.** From time to time ELAP and ELAR run special workshops on aspects of language documentation that draw upon the knowledge and skills of specialist instructors. We have had two such workshops to date: one dealing with audio recording, digitization and archiving, and another with extensible markup language (XML) and its role in language documentation research<sup>4</sup>. The following is an outline of the two workshops:

<sup>4</sup> XML is a document description language, used to describe the content of structured documents – each part of a structured document is described within a defined and logical structure (the structure can be documented in an XML schema or “DTD”). XML documents can be designed, created, processed and transformed manually or by using editors, stylesheets (XSLT “extensible stylesheet language for transformations”), and document processing scripts. It is the preferred data format for text materials, especially for archiving, and a number of language documentation tools, such as Transcriber and ELAN, store data in XML format.

**1. Audio Recording, Digitization and Archiving.** Presented by Professor Dietrich Schueler, Phonogrammarchiv, Austrian Academy of Sciences

TABLE 6. Audio Recording, Digitization and Archiving workshop

	Session	Topics
11 am - 1pm	Principles of audio conversion and preservation (IASA-TC 03)	analogue-digital conversion, choosing digital resolution and file formats, digital to digital ingest
	Optimizing signal extraction from magnetic tape	identifying and assessing track formats and misaligned recording heads, historical equalizations
2pm – 4pm	Field audio recording	microphone arrays, including psychoacoustic considerations, recording devices, post R-DAT
	Small scale digital archiving	manual approach to digital archiving, archiving in the field

**2. XML Day.** This workshop consisted of a mix of tutorial, practical, and discussion sessions, including group work and problem solving. It was designed to provide participants with opportunities to learn about the history, purpose, and formalism of XML, understand its applications, strengths, and weaknesses, see how XML can be applied to linguistic data, gain basic hands on experience with designing, “reading”, evaluating, and editing XML, and learn about technologies that are closely related to and used in conjunction with XML. The schedule for this workshop is given in the following table:

TABLE 7. XML workshop

10:00 - 10:40	Roots, principles and formalism. Look-ahead to architecture. Namespaces and Unicode. Workflow. Discussion of preparatory material.
10:40 - 11:20	Introduction to Oxygen. Practical exercises on document marking up and well-formedness
11:20 - 11:40	Break

11:40 - 12:40	Constraining document grammar: DTD and Schema. Constructing DTDs. XPath and XQuery.
12:40 - 13:00	More Oxygen functions. XPath exercises.
13:00 - 14:00	Lunch break (lunch not supplied)
14:00 - 14:40	XSLT and CSS, XSL:FO
14:40 - 15:20	XSLT exercises
15:20 - 15:40	Break
15:40 - 16:20	Real world: office applications and creating and manipulating XML
16:20 - 17:00	Exercises: creating and manipulating XML, bringing it all together

The content covered in the workshop included theory, practice and architecture, with the following specific topics addressed:

**Theory** XML roots, principles and formalism; XML vs. relational databases

**Practice** Case studies in representing linguistic data and theories; good XML design; XML editors; mobilization; born-XML and marriages of convenience; examples of XML in linguistics: modelling interlinear glossed text, Transcriber, ELAN files

**Architecture** Constraining XML with DTDs and Schemas; namespaces and different XML vocabularies; transforming/displaying XML with XSLT, CSS, and XSL-FO; searching XML using XPath and XQuery; Unicode and XML

**6. SUCCESSES AND CHALLENGES.** Although ELAP has only been in operation since 2003 (and ELAR since 2004) we can point to some successes in our training programs. These can be summarized as follows:

- several students who completed the MA have gone on to work in local language centers (in Australia, and Alaska), and one has recently been offered a job with an NGO in India that deals with cultural and linguistic conservation and revitalization;
- 13 students who completed the MA went on to enrol in the PhD, and all have been awarded scholarships or bursaries;



- all our PhD students have successfully applied for competitive research grants to fund their fieldwork. This fieldwork has been carried out in Iran, Mexico, Vanuatu, Senegal, Nigeria, the Solomon Islands and India;
- our current PhDs include students from Australia, India, Senegal, Sierra Leone, Sweden, Germany, Italy, and UK. Our MA students come from a wide range of countries, including US, Canada, and Australia, as well as the EU and UK;
- we have four post-doctoral fellows (two externally funded) and many visitors who together create a lively research environment for postgraduate students;
- the ELDP training courses have been well received and attain positive evaluations from attendees who have found them useful for their project work.

This success has also been accompanied by some challenges. The following are some of our current major concerns:

A lack of linguistic background for 2/3 of our MA students means that they have to work very hard to acquire basic linguistics knowledge and analytical skills at a very fast pace. We are expected to take students from a range of backgrounds, and it is usual in the UK for MA courses not to require previous knowledge of the content area. We have recently addressed this concern by expanding our Principles of Linguistic Analysis course to two terms (20 weeks) and creating two pathways in the MA degree so that the Field methods course in particular is only taken by students with sufficient linguistic knowledge.

The Field methods course is difficult to organize for the full MA cohort of 17 students to ensure that there are enough small group sessions with the language consultant. Starting in 2007-08, the number of students taking Field methods will be reduced due to the introduction of the pathways, however we may still have difficulties with this course component.

Designing and implementing courses while supervising an average of 17 MAs and 12 PhDs each year has placed a heavy workload burden on ELAP which has only two permanent staff and two post-doctoral fellows. We have received some assistance from ELAR staff who have taught course components and run specialist workshops, however the 'start-up costs' of getting the postgraduate program going have been high. The number of staff in ELAP increases to three permanent staff, one 3-year research fellow, and two post-doctoral fellows from September 2007, and this may relieve some of the work pressures being felt by ELAP staff.

We have not been as successful with PhD student recruitment as we would have liked, especially in attracting native speakers from third world countries (we currently have only two native-speaker students from Africa).

It has been difficult to obtain scholarship funding, especially for MA students. Currently SOAS charges £3,500 (approximately \$US 7,000) in fees for EU and UK students, and £10,500 (approximately \$US 21,000) for others; this level of cost can be difficult for students to pay, along with living costs in London, which are quite high.

We have had success in gaining fieldwork funding for PhD students but this is not guaranteed in the future. An additional problem is locating safe and secure field sites for students to work in. Already, one female PhD student has had difficulties with her field site in southern Nigeria, while another lived through an earthquake and tsunami in the Solomon Islands. One PhD project planned for India was cancelled due to Maoist guerrilla attacks, and we must exercise a high level of care in attending to the risks associated with fieldwork.

There is pressure from the university and the UK government for PhD students to complete their degrees within 3 to 3.5 years, however this time frame is not realistic when fieldwork and corpus development needs to be included, especially for students working in difficult field sites with previously unrecorded languages. So far, the first two of our students enrolled in the PhD are both in their fourth year and will need this amount of time to ensure submission of a dissertation of sufficient quality.

It is difficult at this point in time to fully evaluate the outcomes of our postgraduate courses and training programs. Documentary linguistics is a very new enterprise and our MA and PhD courses are unique in the world as being specialist programs designed to develop and implement theory and practice of documentary linguistics.

The future development of career paths for graduates from our MA and PhD are not yet clear. One PhD student who has almost completed has been awarded a post-doctoral fellowship in Mexico that he plans to take up in 2008, however the career paths of all our PhD students will only emerge over the next few years.

Practitioners of documentary linguistics need to improve their communication with the linguistic community and the wider world. There continues to be a level of ignorance about (and sometimes opposition to) the goals, theory and methods of language documentation among the academic linguistics community and a low level of appreciation among the general community about the value of our work. ELAP staff and students have made numerous presentations in public events to describe our work, however more needs to be done, especially to ensure continued support and funding for the documentary linguistics enterprise.

**7. CONCLUSIONS.** Documentary linguistics is a relatively new field and the practice of language documentation has really only begun in earnest in the past five or so years. We are yet to see the range of possible outcomes from this new approach, or the impact that it will have on the academic field of linguistics, on the communities of speakers of endangered languages, and on the wider world in general. Our experience so far at SOAS however indicates that the training of current researchers and a new generation of language documentation specialists will be both exciting and rewarding for all concerned, and not lacking in challenges.

## REFERENCES

- AUSTIN, PETER K. 2006. Data and language documentation. In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Fundamentals of Language Documentation*. Berlin: Mouton de Gruyter, 87-112.
- AUSTIN, PETER K. 2007. 'Survival of Languages' in Emily F. Shuckburgh (ed.) *Survival: Darwin College Lectures*. Cambridge: Cambridge University Press.
- AUSTIN, PETER K. and LENORE A. GRENOBLE. 2007. Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 12-25. London: SOAS.
- BIRD, STEPHEN and GARY SIMONS. 2003. Seven dimensions of portability for language documentation and description. *Language* 79/3:557-82.
- CSATO, EVA A. and DAVID NATHAN. 2003. Multimedia and documentation of endangered languages. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 1, 73-84. London: SOAS.
- EVANS, NICHOLAS and HANS-JURGEN SASSE. 2007. Searching for meaning in the Library of Babel: field semantics and problems of digital archiving. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 58-99. London: SOAS.
- GIPPERT, JOST, NIKOLAUS P. HIMMELMAN and ULRIKE MOSEL (eds.). 2006. *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178.)* Berlin: Mouton de Gruyter.
- GRINEVALD, COLETTE. 2003. Speakers and documentation of endangered languages. In Peter K Austin (ed.) *Language Documentation and Description Volume 1*, 52-72. London: SOAS.
- HANS RAISING ENDANGERED LANGUAGES PROJECT. 2006. What is language documentation? Available at <http://www.hrelp.org/documentation/whatisit/>, accessed 28th May 2007.
- HIMMELMAN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-95.
- HIMMELMANN, NIKOLAUS P. 2002. Documentary and descriptive linguistics (full version). In Osamu Sakiyama and Fubito Endo (eds.) *Lectures on Endangered Languages: 5*, 37-83. Kyoto: Endangered Languages of the Pacific Rim.
- HIMMELMANN, NIKOLAUS P. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) 2006. *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178.)*, 1-30. Berlin: Mouton de Gruyter.
- MUNRO, ROBERT. 2005. The digital skills of language documentation In Peter K. Austin (ed.) *Language Documentation and Description Volume 3*, 141-156. London: SOAS
- NATHAN, DAVID. 2006. Thick interfaces: Mobilizing language documentation with multimedia. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation (Trends in Linguistics. Studies and Monographs, 178)*, 363-379. Berlin: Mouton de Gruyter.
- NATHAN, DAVID and PETER K. AUSTIN. 2004. Reconceiving metadata: language documentation through thick and thin. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol. 2, 179-187. London: School of Oriental and African Studies.
- NATHAN, DAVID and EVA CSATO. 2006. Multimedia: a community-oriented information and

communication technology. In Anju Saxena and Lars Borin (eds.) *Lesser-known languages of South Asia*, 257-277. Berlin: Mouton de Gruyter.

THIEBERGER, NICK and SIMON MUSGRAVE. 2007. Documentary linguistics and ethical issues. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 26-37. London: SOAS

WALCOTT, HARRY F. 1999 *Ethnography: a way of seeing*. Walnut Creek: Alta Mira.

WHALEN, DOUG. 2003. How the study of endangered languages will revolutionize linguistics. *International Congress of Linguists*, Prague.

WOODBURY, TONY. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language documentation and description*. Volume 1, 35-51. SOAS: University of London.

Peter K. Austin  
pa2@soas.ac.uk

# 4

## SIL International and Endangered Austronesian Languages

J. Stephen Quakenbush

*SIL International*

SIL International has been partnering with Austronesian language communities in language development for over fifty years. This paper briefly reviews that history, situates it in the current environment of international concern for the documentation and revitalization of endangered languages, and looks at ways in which SIL might assist endangered Austronesian language communities of today. Two aspects of *language development* are considered—one more “academic” in nature, focusing on products primarily of interest to linguists and other researchers; the other more “development” in nature, focusing on language resources and competencies of greater interest and relevance to language communities. The paper summarizes some recent studies related to *language endangerment/vitality*, and considers how language development relates to *language revitalization* and *documentary linguistics*. SIL can continue to learn from and link with others in *describing and documenting* endangered Austronesian languages, in providing *consulting and training* at the request of language communities and others, and in designing and developing affordable *language software* to help accomplish related tasks.

**1. INTRODUCTION**<sup>1</sup>. Five thousand people are gathered outside a rural conference center in Kota Belud, Sabah, Malaysia to hear government officials, educators, and representatives of international organizations and local language communities address the role of local languages in a multilingual country. The event is the first Malaysian Indigenous Peoples’ Conference on Education (MIPCE), a week-long conference organized by three Malaysian language community organizations (Kebudayaan Iranun, Kadazandusun Language Foundation, and United Sabah Bajau Organization), together with UNESCO and SIL International.<sup>2</sup> After the fanfare of the opening day, a group of 155 participants gets to work. This smaller group includes representatives from twenty-three language communities in Malaysia, as well as Maori-speaking guest presenters from New Zealand, and representatives from two language communities of Indonesia. The participants are uniformly interested in revitalizing and developing their heritage languages in a national context that officially promotes unity through diversity, but of necessity has focused primarily on the

---

<sup>1</sup> I gratefully acknowledge the assistance of numerous SIL colleagues in the preparation of this paper, including: David Bosma, Mike Cahill, Nico Daams, Diane Dekker, Steve Echerd, Jim Ellis, Paul Frank, Joe Grimes, Tom Headland, Sue Hasselbring, Lou Hohulin, Greg Huteson, Larry Jones, Kimmo Kosonen, Lynn Landweer, Paul Lewis, Rob Lovatt, Rundell Maree, Norm McGuire, Susan Malone, Chip Sanders, Gary Simons, Jim and Karla Smith, Neville Southwell, Barbara Trudell, Bob Weber and Catherine Young. Each provided valuable information and/or critique of specific components. I also thank two anonymous reviewers for their helpful comments. Shortcomings remain my own responsibility.

<sup>2</sup> For more information on MIPCE, held 23-30 April 2007, see <http://www.klf.com.my/mipce1/index.htm>.

promotion and development of a national language. They are especially focused on the role of local languages in the formal educational system. Interest is high, as well as the apparent level of commitment to action. These community leaders and representatives have looked into the future, and have seen that the welfare and very survival of their languages and cultures may be influenced by commitments made and actions taken at this conference.

The first Malaysian Indigenous Peoples Conference on Education illustrates several social and political realities of the global context for Austronesian language communities, endangered or otherwise. Austronesian language communities of today are generally multilingual, linked to the outside world and increasingly networked with each other, ready to learn from the experience of others, and at varying levels of commitment and readiness to promote and develop their own languages.<sup>3</sup> This paper briefly considers the history of SIL involvement in Austronesian languages, and looks at ways in which SIL may best contribute further to efforts on behalf of endangered Austronesian languages.

**2. WHAT IS SIL INTERNATIONAL?** SIL International is a faith-based, international non-governmental organization that partners with language communities worldwide in language development-related efforts. SIL is “faith-based” in that its personnel share a Christian commitment of service to God through service among minority language communities of the world. It is “international” in that members come from over sixty different countries, and carry out language research in over seventy. It is “non-governmental” in that it consists of volunteers serving as private individuals in a non-sectarian, non-profit organization dedicated to a professional, scholarly, community-based approach to language development. “Language development” includes a variety of activities and products, some of which are more easily classified as “academic” and others which more clearly fall into the “(community) development” category. Academic efforts include the publication of scholarly descriptions of languages. Development efforts include the production of literacy materials in local languages, as well as the introduction of literacy itself in communities where the local language may have previously been unwritten.

The hallmark of SIL has traditionally been fieldwork carried out by personnel residing for extended periods of time in local language communities,<sup>4</sup> and specializing in the application of linguistic research to literacy and translation needs. This type of fieldwork inherently involves some level of commitment to “language documentation”<sup>5</sup> as well as to building capacity in the language community to carry out language development efforts. For those concerned with the documentation and revitalization of endangered Austronesian

---

<sup>3</sup> I owe the basic articulation of these characteristics to an unpublished paper by John Watters entitled “Minority language communities in the 21<sup>st</sup> century.”

<sup>4</sup> These “extended periods of time” often span decades, with an SIL team resident or near to the community much of that time.

<sup>5</sup> The concept of “language documentation” as the product of “documentary linguistics” is discussed further below. The primary distinction of language “documentation” is its focus on primary data, collected, annotated, and made available as “a lasting multipurpose record of a language” (cf. Himmelmann 2006:1).

languages, then, it may be instructive to review SIL's experience in Austronesian languages, and to consider ways in which SIL might further assist those working for the well-being of Austronesian languages and their speakers.

**2.1 A BRIEF HISTORY.** SIL began in the 1930s, when the scientific description of non-Indo-European languages was in its infancy. Cameron Townsend held the first "summer institute of linguistics" in Sulphur Springs, Arkansas in 1934, and soon thereafter led a small group of North American linguists to serve among indigenous language communities of Mexico. From the start, Townsend committed to rendering practical assistance to language communities (such as training in alternative gardening techniques) as well as informal advocacy at the local and national government levels. Practical assistance and informal advocacy was offered alongside a program of formal research, which in turn shaped local literacy and translation programs. Townsend was also committed to serving the national government. He became a close friend of Mexican President Lázaro Cárdenas, and was so impressed with the President's commitment to the welfare of indigenous peoples that he eventually wrote a biography highlighting the President's accomplishments (Townsend 1952).

This model of cooperation with governments in service to indigenous peoples was transplanted to the realm of Austronesian languages in 1953, when Dr. Richard S. Pittman arrived in the Philippines with a small group of linguists, at the invitation of future President Ramon Magsaysay. Beginning in the Philippines, the work of SIL spread to Papua New Guinea (1956), Indonesia (1972), and Malaysia (1977).<sup>6</sup> SIL officially began assisting in language development efforts in Rapa Nui of Easter Island in 1976.<sup>7</sup> Through the 1980s, SIL began serving various language communities in the Solomon Islands, New Caledonia and Vanuatu.

**3. WHAT IS LANGUAGE DEVELOPMENT?** Language development, as noted above, includes some activities and products that are more purely academic in nature relating to language description, and others that are more clearly related to the overall welfare and development of language communities.<sup>8</sup> From the beginning, SIL academic goals for language projects have included the publication of a grammar, some type of lexicon and a collection of text material. A standard grammar sketch also includes a basic phonology statement. The lexicon may range from a simple glossary or word list to more elaborate dictionaries with thousands of entries containing detailed grammatical information and sample sentences. A glossed and annotated text collection provides at least representa-

---

<sup>6</sup> Papua New Guinea and the Indonesian province of Papua, being more linguistically complex areas, also represent many non-Austronesian languages. Approximately one fourth of Papua New Guinea's 820 living languages are Austronesian in origin, according to Ross's (1988) estimate of 201 Austronesian languages.

<sup>7</sup> See Programa Rapa Nui (2006) for an extensive bibliography of publications in and on the Rapa Nui language. Weber and Weber (1998) report on early educational efforts in that language development program.

<sup>8</sup> Over the decades, SIL has become increasingly aware of and committed to the essential role the language community plays in determining its own goals and needs in language development work.

tive data on which the phonology, grammar and lexicon are based, as well as documenting important cultural information. Some text collections provide more detailed linguistic analysis through morpheme-by-morpheme glosses, while others show less linguistic detail but provide more raw material for its anthropological content. These three basic tools of a grammar, lexicon and text collection have long been considered foundational for language description as an academic undertaking, as well as for language development work in general.

Also from the beginning, SIL goals have included activities and products that more directly relate to the overall well-being and development of the language community itself. Primary among these goals has been the promotion of literacy in the local language. Literacy in any language, of course, presupposes the development of an orthography and literature in that language. Where no orthography exists, SIL works with the community to develop one, taking into account linguistic, pedagogical, political and other factors. The initial development of reading materials often happens through writers' workshops.<sup>9</sup> Materials for teaching reading are designed taking into account whether the learners will be learning to read for the first time or are already literate in another language. In many cases, community literacy efforts have taken the form of adult, non-formal education campaigns. Increasingly, in today's interconnected world, promotion of literacy has focused more on the role of the local languages in formal, multilingual education—especially in the earlier levels of formal schooling. With a growing awareness on the part of educators, policy makers, and multilateral agencies regarding the crucial role of language in education, SIL is increasingly being called on by language communities as well as national governments and other agencies to advise on matters of multilingual education, and to advocate on behalf of minority language communities.

Language development, then, may be defined as the advancement of language resources and competencies so that a community can effectively use its language(s) for the varied purposes it requires and desires. While it is impossible to draw a clear line between resources and competencies, *language resources* primarily refers to published products that facilitate and “legitimize” the use of a language for specific purposes, in the eyes of its own speakers as well as in the estimation of other significant opinion leaders and decision makers. These products include such items as grammars, dictionaries, lists of specialized vocabulary, published text materials, etc. *Language competencies*, on the other hand, refers to particular skills in the language—literacy being primary among these. Other skills would involve use of the language for public (including electronic) discourse and artistic expression in a variety of media. Such language competencies can be measured on an individual level. But competency in *language development* also involves a community's capacity to produce language resources as well as to promote specific skills in and uses for a language among the members of that community.

---

<sup>9</sup> See Weber, Wroge and Yoder (2007) for a discussion on the place of writers' workshops in the context of a community literacy program.



**4. SIL'S EXPERIENCE IN AUSTRONESIAN LANGUAGES.** *Ethnologue* lists 1,268 Austronesian languages, making it the second largest language family in the world.<sup>10</sup> (Only the Niger-Congo family has more languages—at 1,514.) SIL has had some level of involvement in 393, or roughly one third of all Austronesian languages. That history of involvement is summarized here in terms of publications and other types of activities, before proceeding to look at the issue of “endangered” languages in particular.

Publishing a description or undertaking any type of project in a given language presupposes, of course, that such a language has been identified. The question of what or how many Austronesian languages are in existence is by no means a simple one. That *Ethnologue* can list 1,268 of them indicates a tremendous amount of resources invested in language survey. One of the services SIL has provided the governments of the Philippines, Malaysia, Indonesia and Papua New Guinea has been in the identification of languages spoken in the more remote areas of those countries. While Austronesian linguistics has been the purview of numerous scholars and institutions around the region and the world, it is not an exaggeration to state that a considerable amount of what is known about the identity and interrelationships of Austronesian languages can be linked to data gathered by SIL field personnel.<sup>11</sup> In addition to data collected by SIL researchers, SIL attempts to track and report basic information on all the world's languages through *Ethnologue*.

The clearest indicator of a significant level of involvement in a particular language is a record of publication in or on that language. Figure 1 shows the total number of SIL publications on Austronesian languages, classified according to the relative size of the language group. Publications are subcategorized according to whether they are primarily of an academic nature (typically in a language of wider communication), or in the vernacular (meant for the local language community). The general picture is one of academic publications slightly outnumbering vernacular publications, and with SIL involvement increasing along with the number of languages in a particular size category.

A record of publications alone does not show the extent of SIL involvement with a given language community, however, either in terms of time depth or scope of activities. In some cases, as with several languages in the southern Philippines (Tboli, Blaan, Western Subanon, Binukid and some varieties of Manobo, for example), SIL and Translators Association of the Philippines assisted communities in developing highly successful programs to introduce and promote literacy in the local language.<sup>12</sup> Such efforts have necessarily involved the standardization of orthographies, production of literacy materials, and equipping members of the language communities as literacy teachers, often over a period of decades.

In many cases, especially where government schools are already operable, SIL has been less active in literacy and education efforts. One notable exception is in the Lubuagan, Kalinga community of northern Luzon, also in the Philippines. The municipality of Lubuagan has since 1998 been the site of a mother tongue multilingual education program

<sup>10</sup> See [http://www.ethnologue.com/show\\_family.asp?subid=89851](http://www.ethnologue.com/show_family.asp?subid=89851)

<sup>11</sup> For some Philippine examples of SIL contributions, see Walton 1979 and Gallman 1997, as well as acknowledgements in such works as Zorc 1977 and McFarland 1981.

<sup>12</sup> See Awid 2001 for a discussion of an innovative literacy program in Western Subanon.

known as the “First Language Component/Bridging Program”.<sup>13</sup> The goals of this program are to equip elementary school teachers to design and produce their own language resource materials and to use the first language of the students effectively for the introduction of literacy skills. Use of the local language is emphasized especially in the first three grades, bridging into increased use of Filipino and English for both literacy education and delivery of curriculum content in a sequential manner in succeeding years. Initial indicators demonstrate the effectiveness of such an approach, as measured by student performance on standardized tests in other academic subjects as well as in language arts.<sup>14</sup> The Lubuagan FLC Bridging Program is being documented carefully over a period of ten years, as part of a larger Longitudinal Studies Project, which is also investigating similar programs in mainland Asia and West Africa.

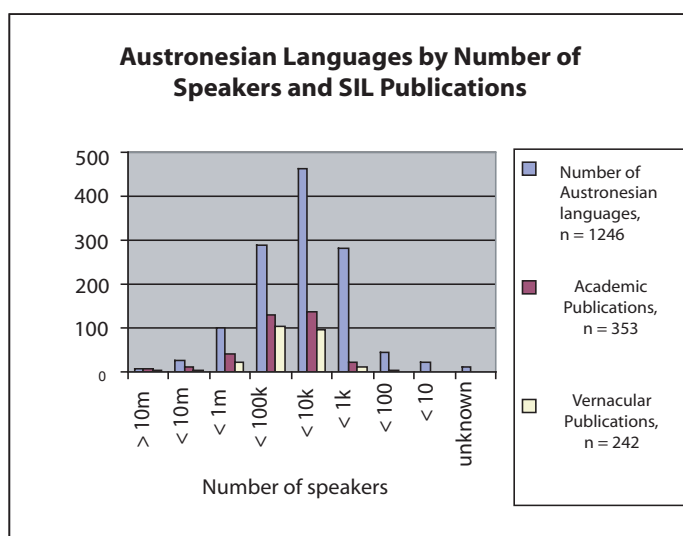


FIGURE 1. Austronesian languages and SIL publications<sup>15</sup>

The country in the Asia-Pacific region where SIL has been most active in promoting literacy and education in the local language has been Papua New Guinea, which has made great strides in incorporating the local language, or *tokples*, into the formal school system and related pre-school efforts. Papua New Guinea provides numerous case studies for the

<sup>13</sup> For further information on the Lubuagan program, see especially Dekker and Dumatog 2003. See also Dekker 1999, Young 2002 and Dekker and Young 2005.

<sup>14</sup> A January 5, 2007 letter from the Department of Education (Division of Kalinga) attributes Lubuagan’s high scores on a National Reading Test in English and Filipino to the First Language Component. Lubuagan ranked first among the ten districts where the test was administered.

<sup>15</sup> I am indebted to Gary Simons for these figures, which come from SIL’s Language Program Database, as accessed in February 2007.

role of language communities in developing their own pre-school programs in the local languages, as well as for an evolving national language policy increasingly supportive to the development and use of local languages.<sup>16</sup>

Austronesian languages have been considered thus far in this paper without reference to whether they should be characterized as endangered. In order to address the issue of SIL and endangered Austronesian languages, we must first come to some agreement on what is meant by “endangered.”

**5. WHAT IS AN “ENDANGERED LANGUAGE”?** Numerous terminologies have been suggested for classifying degrees of language endangerment, with little real agreement among linguists regarding a standard. The crudest measure of language endangerment would be in terms of absolute number of speakers. It has often been observed, however, that the number of speakers which make a language seem threatened varies according to geography. In Africa, where the median language size is over 25,000 speakers, a language of 1,000 speakers may in fact be quite endangered. In the Pacific, however, where the median language size is only 800, languages even smaller than that may remain vital for several generations to come, depending on other variables.<sup>17</sup>

Fishman’s (1991) “Graded Intergenerational Disruption Scale” is one of the better known systems for classification of levels of language endangerment. The key indicator in this scale remains whether languages are being passed on intergenerationally. Table 1 below summarizes the eight stages, along with suggestions for “reversing language shift” in stages 4 through 8. The vast majority of Austronesian languages would fall between stages 5 and 8 on the GIDS, with 5 being the more secure ranking, and eight being the most endangered. Very few languages of the world reach level 1.

TABLE 1. Fishman’s (1991: 87-111) Graded Intergenerational Disruption Scale (GIDS) for Threatened Languages (format taken from Baker 2006: 61).

Stage 8	Social isolation of the few remaining speakers of the minority language. Need to record the language for later possible reconstruction
Stage 7	Minority language used by older and not younger generation. Need to multiply the language in the younger generation
Stage 6	Minority language is passed on from generation to generation and is used in the community. Need to support the family in intergenerational continuity (e.g. provision of minority language nursery schools)

<sup>16</sup> See Litteral (1999 and 2004) and Wroge 2002 for some helpful summaries of language and education issues in Papua New Guinea. See [http://www.sil.org/literacy/snapshots/snap\\_STEP.htm](http://www.sil.org/literacy/snapshots/snap_STEP.htm) for information regarding an SIL-sponsored training program for teachers in mother tongue education and literacy programs.

<sup>17</sup> Barbara F. Grimes 2001 and Joseph E. Grimes 1995 discuss various factors relating to language viability in addition to population size.

Stage 5	Literacy in the minority language. Need to support literacy movements in the minority language, particularly when there is no government support.
Stage 4	Formal, compulsory education available in the minority language. May need to be financially supported by the minority language community.
Stage 3	Use of the minority language in less specialised work areas involving interaction with majority language speakers.
Stage 2	Lower government services and mass media available in the minority language.
Stage 1	Some use of the minority language available in higher education, central government and the national media.

A framework for categorizing language endangerment developed by a UNESCO team of experts uses the following nine factors:

- a. Intergenerational language transmission
- b. Absolute numbers of speakers
- c. Proportion of speakers within the total population
- d. Loss of existing language domains
- e. Response to new domains and media
- f. Materials for language education and literacy
- g. Governmental and institutional language attitudes and policies
- h. Community members’ attitudes towards their own language
- i. Amount and quality of documentation (Brenzinger 2003: 9-17)

Each of these categories (except for #2, which is a raw number) receives a ranking on a scale of 0 to 5, with 5 being the most robust evaluation. Lewis (2006) attempts to apply this framework to a hundred languages of the world. He concludes that although it represents an admirable attempt to capture “state of the art” thinking on language endangerment, practical use of the framework is limited because so little is currently known about many of the languages of the world, especially those which are most endangered.

The flip side of language endangerment is language vitality, and some scholars have chosen to view matters from this more positive perspective. Landweer (2006: 65-68) proposes the following eight Indicators of Ethnolinguistic Vitality. Each of these indicators is assigned a point value from 0 to 3, with 3 being the value that will lead to the most robust evaluation.

1. Position of the speech community on the remote-urban continuum  
(less frequent contact is better)
2. Domains in which the target language is used  
(more domains for the vernacular is better)
3. Frequency and type of code switching  
(less code switching is better)

4. Population and growth dynamics  
(more population is better)
5. Distribution of speakers within their own social network  
(tighter social structure for the vernacular is better)
6. Social outlook regarding and within the speech community  
(higher group prestige is better)
7. Language prestige  
(higher language prestige is better)
8. Access to a stable and acceptable economic base  
(more stable and acceptable income is better)

Landweer's more nuanced system grew out of research on language endangerment studies from other parts of the world, coupled with long experience in the context of Papua New Guinea. She argues that the Melanesian context is distinct from much of the rest of the world, due to the relative lack of a dominant "oppressor" language and "the traditional egalitarian coexistence of multiple languages in Pacific nations" (2006: 5). It may be that Landweer's system is more appropriate for the evaluation of most contexts where Austronesian languages are spoken, especially in the absence of a clearly dominating cultural group and language. Still, it is apparent that relatively in-depth information on each language group must be available in order to make the best use of the indicators.

Although the classifications for degrees of endangerment are several and varied, there is one basic underlying principle: a language is moving in a dangerous direction if it is losing ground in terms of number of speakers or number of contexts in which it is used. *Ethnologue* lists 516 "Nearly Extinct Languages" around the world, where "only a few elderly speakers are living."<sup>18</sup> The two regions with the most languages in this category are the Americas (170 out of 1002 living languages=17%) and the Pacific (210 out of 1345 living languages=16%). The high number of nearly extinct languages in the Pacific might at first seem to belie Landweer's claim that Melanesian languages are by and large vital ones. However, 168 (or 61%) of the 210 nearly extinct languages in the Pacific are not Austronesian languages, but rather Aboriginal languages of Australia. Excluding Australia, the percentage of nearly extinct languages in the rest of the Pacific is only 4%. Figure 2 shows the percentage of living languages that are nearly extinct according to region of the world, with Australia listed as its own region.

There are two basic reasons why a language loses speakers. One is that the community of speakers of the language are physically dying out. The other is that people shift to using another language in place of the first. Throughout history, some languages have disappeared due to entire populations dying out as the result of disease, war, or natural calamity. Cahill (1999) discusses the situations of language groups of a few hundred speakers or less in Brazil and Papua New Guinea that were in imminent danger of dying out completely. Each has made a comeback, thanks in part to access to modern medicine. An important

<sup>18</sup> See [http://www.ethnologue.com/nearly\\_extinct.asp#Africa](http://www.ethnologue.com/nearly_extinct.asp#Africa).

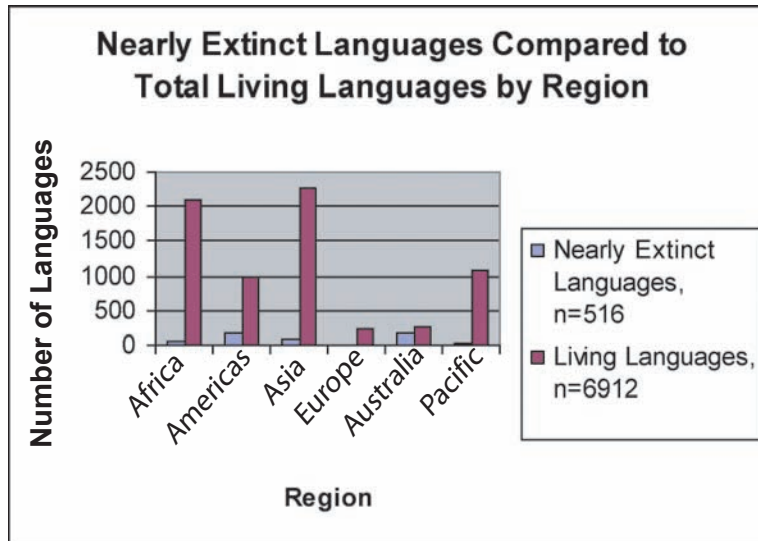


FIGURE 2. Nearly extinct languages compared to living languages according to region of the world.

factor in several of these cases was also a heightened sense of self-respect among language community members, corresponding with the introduction of literacy and the translation of the Bible in these languages.

In most cases, language death is not so much about an entire population dying out, as it is the end point of language shift—the cumulative result of countless individual choices of a population to no longer use their heritage language. Language death typically happens gradually, as more and more parents decide to speak to their children in a national or regional language because they believe their children would be better served by speaking that language well. Thus more and more children begin learning a national or regional language at home to the exclusion of their heritage language. This dynamic of endangerment has no doubt been playing out in the twenty-seven nearly extinct Austronesian languages listed in *Ethnologue*, as well as in an uncounted number of other Austronesian languages experiencing varying levels of endangerment.<sup>19</sup>

What factors will work most strongly in favor of language maintenance as opposed to shift? Landweer found through her in-depth case studies in Papua New Guinea that two primary social factors related positively to language vitality—language as a marker of ethnicity, and endogamous marriage patterns. A language's vitality is more at risk when the link between language and ethnic identity weakens, and in situations where there are an increasing number of exogamous marriages.

<sup>19</sup> For example, Bob Weber (personal communication) reports that in spite of numerous language preservation efforts on the part of school teachers and others, Rapa Nui of Easter Island is increasingly endangered due to lack of intergenerational transmission.

Scebold (2003) describes a situation in Central Tagbanwa of Palawan, Philippines where only the older generation speak the Tagbanwa language to any real degree of fluency. Over the decades since World War II, settlement patterns, disease, and massive in-migrations of speakers of other languages influenced Tagbanwa parents to switch to the use of Cuyonon, a regional lingua franca, with their own children. The result is a younger generation that maintains relatively few markers of Tagbanwa ethnicity, and a heritage language that is on the brink of extinction.

Headland (2003) expands the definition of endangerment in his discussion of thirty-two Negrito languages of the Philippines. Although some of these languages are now extinct because all their speakers have died out (or sometimes been killed), he argues that most of these languages are endangered due to a rapid degree of culture change, resulting in loss of broad areas of traditional vocabulary. Because the social and cultural contexts of these language communities are undergoing such sweeping changes, the languages themselves are becoming unrecognizable compared to earlier forms of the languages reflecting the world and worldview of hunter-gatherer societies. Headland's "endangerment" then, focuses not on languages per se, but on the breadth and richness of the lexicon.

In the final analysis, distinguishing levels of endangerment or vitality is probably less important than encouraging a perspective of linguistic diversity as an asset as opposed to a liability, and granting that communities of human beings have an inherent right to appreciate, promote, maintain and develop their unique languages as resources for communication and expression of their unique cultural heritage and identities.<sup>20</sup> While maintaining linguistic diversity for its own sake has not been an explicit or primary goal for SIL in the past, the academic research and development activities in which it engages do provide essential resources for the well-being of endangered languages.<sup>21</sup>

**6. LANGUAGE DEVELOPMENT AS LANGUAGE REVITALIZATION.** SIL serves language communities of the world in development-related efforts through research, translation and literacy. Although SIL serves primarily in smaller language communities, the issue of the survival of these languages has largely been assumed, rather than addressed.<sup>22</sup> This is perhaps even more true in the case of Austronesian languages, where endangerment has not yet reached the same alarming levels as in Australia and the Americas. Lack of awareness of language endangerment issues on the part of SIL and lack of advocacy on behalf of minority ethnolinguistic groups represent weaknesses to be addressed. Happily, however, the traditional goals of SIL projects do contribute positively and crucially to efforts on behalf of endangered languages.

Language development has been defined above as the advancement of language resources and competencies so that a community can use its language(s) for the purposes it requires and desires. Language endangerment was summarized as a situation where a language was "losing ground" in terms of number of speakers or number of contexts in

---

<sup>20</sup> See Grin 2005 for an interesting discussion of human rights and other arguments in favor of linguistic diversity.

<sup>21</sup> See Pittman 1998 for a collection of anecdotes documenting "sixteen accounts of ethnic renaissance" in language communities where SIL has had an active presence.

<sup>22</sup> See Headland 2003 and Scebold 2003 for recent exceptions with regard to Philippine languages.

which it is used. In the most severe cases, a population of speakers is either actually dying out, or shifting to another language as the older generation chooses not to use the heritage language with the younger generation. Language revitalization focuses then on “regaining ground,” moving a language from more to less endangered, from less to more vital. More specifically, language revitalization efforts aim to increase the number of speakers or number of contexts in which a language is used.

Clearly, there are common ingredients for successful language development and language revitalization efforts. These include *critical resources*, *critical expertise*, *critical mass*, and a *critical context*. *Critical resources* for language revitalization depend and build on the linguistic analysis and language resource materials that are at the heart of language development. Linguistic analysis provides the necessary foundation for designing orthographies, writing grammars, compiling dictionaries, and publishing text materials.

*Critical expertise* for language development involves basic language competencies on the part of individual members of a language community (such as listening, speaking, reading and writing in the language), as well as competency on the part of the community overall to produce necessary language resources and to promote the use of a language in that community. This kind of capacity presumes a level of proficiency in linguistic analysis within the community, but also requires foundational skills in community organizing as well as in addressing a range of educational, political and promotional concerns.

*Critical mass* refers to the involvement of a significant proportion of the community in language development and revitalization efforts. Experience has shown that it is possible to produce language resources that are never widely used in a community, especially if those products have been generated by a few individuals working in relative isolation from the broader community. Opinion shapers and decision makers throughout the community must be actively involved and supportive in order for language revitalization efforts to succeed.

A *critical context* for language revitalization necessarily assumes a local community committed to the promotion and use of its heritage language. Efforts of the local community, however, must be situated in a larger context of regional and national policies that allow smaller languages to flourish. There is little point in developing written materials for use in a local elementary school, for instance, if national educational policy dictates the use of national and international languages only inside the classroom. Thus, local language communities need not only the permission, but also the support of local and national institutions to use and develop their languages for further use.

Language revitalization is essentially an effort to encourage a community to make certain choices about acquiring and using a certain language over a sustained period of time. The products and activities typical of language development projects can be seen as prerequisite for making these choices possible. This is not to say that producing written materials or introducing the local language as a medium of formal instruction are in every case the most appropriate first steps toward language revitalization.<sup>23</sup> In a case where a language is no longer being passed on to children at home, producing books and school lessons may

---

<sup>23</sup> See Reynard 1999 for a discussion on a “hierarchy of needs based on the health of a language”, which employs Fishman 1991’s Graded Intergenerational Scale of Disruption in his consideration of the language situation in American Indian communities.



be superfluous or even counterproductive. However, Lewis and Trudell (2005: 5) argue—along with many others—that “sustainable language cultivation ultimately requires [that language’s] use in written form by some significant segment of the population.”

Any effort to effect social change requires sustained leadership. Effective, sustainable leadership for language development and revitalization efforts will be community-based, with links to national and international support. Local leaders must ultimately provide the motivation, and display the example and political will to revitalize their languages in their daily lives in the context of their own community and family. Such efforts can be encouraged and practically supported (and sometimes crucially so) by outsiders skilled in language analysis, the production of language resource materials, and other community-building efforts.<sup>24</sup>

**6.1 IBATAN—A REVITALIZED LANGUAGE AND COMMUNITY.** The Ibatan of Babuyan Claro Island in the Northern Philippines exemplifies an endangered Austronesian language community that has been revitalized over the past three decades through sustained language and community development efforts. In 1978<sup>25</sup> there were fewer than 500 speakers of Ibatan, many of whom were no longer using the local language variety in their homes. The more prestigious Ilokano language was preferred for public discourse, and native speakers of Ilokano openly ridiculed the Ibatan language. In the 1980s, SIL personnel and their Ibatan counterparts initiated concerted efforts to promote the use of written and spoken Ibatan in as many contexts as possible. This small team of local language activists and outsider linguists launched a local newspaper, printed numerous books, posted written announcements, started a writers’ club, and formed an economic cooperative—all using Ibatan only. They modeled Ibatan use in school, at church, and in other public contexts. They worked with non-Ibatan government officials to translate resolutions and by-laws into Ibatan, and the Ibatan began writing their own resolutions in the Ibatan language. In all these efforts, the goal was not to prohibit the use of any other language, but to model Ibatan as a viable, valuable alternative for personal and official communication. The publication of a hymn book in 1985, and the New Testament in 1996 further enabled sustained public use of the Ibatan language.

In addition to more purely language-related efforts, the team advocated through the years for the establishment of a high school and medical clinic, and more recently for the granting of a Certificate of Ancestral Domain Title. Prior to the Ibatan submitting their claim, the Philippine government had no official record of the Ibatan as a distinct cultural community. Two factors were crucial in the decision to award the land title: (1) recognition of the Ibatan as a distinct “tribe” (evidenced in large part by the publications in and about

---

<sup>24</sup> See Miller 2001 for a helpful discussion on ways “Mother Tongue Advocates” from the outside can be of assistance to leaders of minority language communities and others.

<sup>25</sup> 1978 was the year SIL linguists Rundell and Judi Maree took up residence on Babuyan Claro Island. They continued to reside among the Ibatan for extended periods of time through 1996, and are still working with Ibatan speakers on a dictionary, text collection and other development projects. A grammar is currently in press.

their distinct language); and (2) acknowledgement of the long-term residence of the Ibatan people on Babuyan Claro Island (evidenced by the genealogical record of generations of Ibatan published by SIL in 2005 at the request of the Ibatan community).<sup>26</sup>

In 2007, there are 1200 to 1300 Ibatan people. They have official recognition from the National Commission of Indigenous Peoples as a distinct tribe of the Philippines, along with a Certificate of Ancestral Domain Title to the land they have historically inhabited. They have a high school and medical clinic on their island. There are a growing number of Ibatan with college degrees. Their traditional arts of reed- and vine-weaving, and making bark cloth are enjoying a modest revival. Ibatan enjoy numerous publications in their own language, and speak it with pride in public as well as at home. And sometimes, even an Ilokano learns a little Ibatan.<sup>27</sup>

**7. DOCUMENTARY LINGUISTICS.** Awareness of the importance of language documentation has been growing worldwide over the past couple of decades along with awareness and concern over language endangerment. Language documentation has to do with producing a lasting record of representative samples of that language. As traditionally practiced by SIL, and indeed by the whole Western linguistic enterprise, language documentation has focused on the production of resources for the linguist or academician more than on resources that directly benefit speakers of the language being documented. In the tradition of early twentieth century American linguists Sapir and Bloomfield, field linguists have produced grammatical descriptions and text collections which would be published by major universities and academic publishing companies in order to advance a general, scholarly understanding of human language.<sup>28</sup>

American structuralist linguistics in general, together with a generation or so of SIL linguists trained in the tagmemic tradition of Kenneth Pike, emphasized the importance of language data (i.e., a record of actually occurring linguistic behavior). This dependence on more or less naturally occurring language data was lost to much of Western linguistics during the Chomskyan era of the latter half of the twentieth century, as linguists in search of language universals increasingly and ironically relied on their own intuition and grammaticality judgments of sample sentences generated by the linguists themselves. Field linguists (those working with more “exotic” languages), on the other hand, remained of necessity more dependent on language data produced by and in a community of native speakers. Although the very enterprise of descriptive linguistics depended on primary language data, the actual data were rarely published. Rather, analytical statements were made on the basis of the data in an attempt to “describe” rather than “document” the language.

Language data that have been published in the tradition of descriptive linguistics have often been “regularized” or “systematized” or even “corrected,” so that the documentation no longer portrays the linguistic behavior as it actually occurred, but rather as a more polished, “limited” product. Some adjustments to the language data were for esthetic reasons,

<sup>26</sup> See Maree 2005.

<sup>27</sup> Rundell Maree (personal communication) reports that at least two Ilokano high school teachers on Babuyan Island have learned enough Ibatan to use the language well in the classroom.

<sup>28</sup> Bloomfield’s 1917 *Tagalog texts with grammatical analysis* is still considered a classic of this sort.

such as the editing out of hesitations or false starts, adjustments in vocabulary, sentence length, etc. These sorts of edits may be made solely by the analyst, but often the original speaker or writer of the language data also desires some degree of editing in order to put their contribution in its best form.<sup>29</sup>

Other adjustments to language data come as an inevitable part of the transcription process. Decisions must be made in transcribing audio-recorded language data, for example, as to what level of detail one wishes to transcribe (broad versus narrow, phonemic versus phonetic). Inevitably, significant segmental and suprasegmental details are lost in the process of committing a spoken text to paper—not to mention the loss of accompanying gestures, props or other important aspects of the communicative context. More rarely, field linguists have video-recorded linguistic behavior, but until very recently, technological considerations have prevented such recordings from being commonly made or widely shared.<sup>30</sup>

In summary, language data on which much linguistic analysis and description is based have rarely been published as such. This is as true of SIL-published data on Austronesian languages as much as it is true of material published by other field linguists on less commonly studied languages around the world. Where language data have been published, it has usually not been “primary data,” but rather “secondary data” that have been edited, systematized or regularized in some way.

The past decade has seen increasing interest in the documentation of representative primary data in a form that will be permanently accessible to speakers and researchers in an electronic environment.<sup>31</sup> Indeed, a new sub-discipline of linguistics has appeared bearing the name of Documentary Linguistics. The web-site of the Hans Rausing Endangered Languages Project at the School of Oriental and African Studies credits Nikolaus Himmelmann as a catalyst for the development of this discipline, citing his 1998 paper entitled “Documentary and descriptive linguistics.” In it, Himmelmann (1998: 116) states that

The aim of a language documentation is to provide a comprehensive record of the linguistic practices characteristic of a given speech community... This... differs fundamentally from... language description [which] aims at the record of a language... as a system of abstract elements, constructions, and rules.

Himmelmann, Gippert and Mosel (2006: v) specify that documentary linguistics is concerned with the “methods, tools and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties.”

---

<sup>29</sup> Himmelmann (2002:16) mentions the “problem” of contributors of language data wanting a “clean, edited” finished product. He suggests a possible compromise solution, where publications of the data in book form (for public use) contain edited versions, while recordings and transcripts of original texts are kept in a database which can be accessed for further scientific inquiry.

<sup>30</sup> Motivation for video recording language data was higher for certain art forms, as with the case of the Manobo oral literature analyzed extensively by Wrigglesworth (1991, 1993, 2004 and In Press).

<sup>31</sup> See Austin 2006 for a fuller discussion on the role and handling of data in language documentation.

Note that we have made an assumption throughout this paper that languages are in fact discrete, whole, abstract yet identifiable and countable systems. Although this assumption matches popular understanding, it is by no means universally accepted in the scholarly world. Himmelmann (2002) successfully avoids the central issue of what a language is by consistently referring to the “linguistic practices of a speech community.” Suffice it to say here that there are several perspectives from which linguistic behavior can be viewed, and that one of these includes the useful fiction that languages exist, and hence can be identified, described, cultivated and propagated.<sup>32</sup>

**8. WHERE TO FROM HERE?** SIL International has been identifying, describing, and assisting in the development of resources in Austronesian languages for over fifty years. SIL activities and goals have, of course, been shaped by more general understandings of what constitutes a language, as well as what constitutes linguistic field work. SIL began with the goal of serving language communities in lasting ways, primarily through engagement with those communities in academic research (especially in descriptive linguistics), translation of literature of high moral and practical value (normally including parts of the Bible), and the promotion of literacy in the local language (sometimes even introducing the concept to a community for the first time). For the past couple of decades, SIL has classified much of what we do as “language development”.<sup>33</sup> Although SIL has stressed service to all—including national governments, universities and other host institutions—in practice much of our work has been focused on local languages at the local community level.<sup>34</sup>

Meanwhile, the international academic world has awakened to the fact that many of today’s languages are endangered—that is, losing ground in terms of number of speakers and/or relative number of domains in which the language is used. Motivated by the belief that a loss of linguistic diversity is a loss to humankind, and that communities have an inherent right to preserve and promote their unique languages and cultures, a growing number of academic, governmental and non-governmental organizations have become committed to the documentation and revitalization of endangered languages. Within the Austronesian language family and beyond, “endangered languages” overlap with the local languages in which SIL has an on-going involvement and active interest.

How can SIL serve local language communities effectively in a manner relevant to the broader context? Fruitful service may best be rendered through continued documentation and description, consulting and training, and making available appropriate, affordable lan-

---

<sup>32</sup> Pike was a proponent both of different perspectives on language-as-system and on explicitly viewing language as a type of human behavior. See Pike1967 for his magnum opus on this foundational aspect of Tagmemics, and Pike 1982 for a more readable summary on the same issues.

<sup>33</sup> Paul Lewis (personal communication) and others of SIL are now promoting the use of the term “language-based development” to describe the holistic, community-based efforts in which SIL engages. This is due in part to terminological confusion related to the use of “language development” by some linguists to refer to child acquisition of language in progressive, developmental fashion.

<sup>34</sup> Focusing on the local community is not bad. It does, however, carry with it the danger that we will fail to notice significant trends or important changes happening in a broader context.

guage software. In each of these spheres, it is important to work collaboratively with local communities, national governments, international agencies, and the academic community at large.

**8.1 DOCUMENTING AND DESCRIBING.** SIL needs to distinguish between documenting and describing languages, and renew its commitment to both as crucial components of language development. Needs and desires of local language communities will increasingly shape the projects in which SIL becomes involved. The end products of language documentation and description will thus serve the local community, as well as national and international bodies. SIL has always been committed to making the results of language research available to all who wish to benefit from them. In this day of computer-assisted research and web-publishing, making language data accessible is increasingly becoming feasible. An indicator of language research becoming more accessible is the fact that the SIL International Publications web page has links to on-line publications specific to languages of the Philippines and Papua New Guinea.<sup>35</sup>

While there is increasing capacity and demand for on-line publishing, language communities will continue to value print publications in the local language of a more traditional or descriptive nature. Multi-language phrase books, word lists, and dictionaries are among the most highly valued and commonly purchased publications that SIL helps produce. In the African context, Batibo (2007: 5) notes a three-fold motivation for those engaged in language documentation—namely, “to preserve data of endangered languages, to build comprehensive linguistic databases and to service the relevant communities.” He states that most linguists in Africa would like to see primary importance given to the last of these three motivations, arguing that language documentation should not just serve to record dying languages, but also to empower communities to use their languages extensively and proactively. He lists the most useful and relevant materials for language empowerment as the following:

1. a practical orthography
2. a user-friendly reference grammar
3. a basic reference dictionary
4. specialized dictionaries or thesauruses of plants, wildlife, environment and indigenous knowledge systems
5. literacy materials (for learning and sustaining literacy)
6. variety of reading materials for cultural and socio-economic empowerment
7. any other relevant materials

Using the Naro language of Botswana as a case study, Batibo notes that the production of these kinds of materials led to empowerment of the language community as manifested by a greater emotional attachment of speakers to their language, higher self esteem among members of the community, greater desire for the language to be used by future generations, and more vigorous efforts on the part of adults to encourage young people to actually

---

<sup>35</sup> <http://www.sil.org/acpub/index.html>

use the language. Batibo's observations indicate that SIL needs to maintain an orientation of practical service to language communities while actively engaging in the new discipline of documentary linguistics.

**8.2 CONSULTING AND TRAINING.** SIL consults with individuals, teams and committees representing specific language communities, as well as with national governments and multilateral agencies. Local language communities, in particular, have changed significantly over the past decades as a result of increasing access to higher formal education and greater integration into their larger socio-political contexts. As a result of these changes, SIL's role has been shifting to a more consultative, collaborative, facilitating one vis-à-vis the local language community. This trend needs to continue. As evidenced at the Malaysian Indigenous People's Conference on Education, local language communities and others are requesting assistance from SIL to advise, advocate for, and link language communities with relevant training and resources. Consulting, training and advocacy efforts range from small, non-formal workshops held in local communities, to larger events in more centralized locations bringing together representatives from several communities, to national level formal training programs, to international academic conferences.

The Northern Philippines Mother Tongue Translators Association (NPMTTA) consists of several local language committees that provide leadership for language development programs, the majority of which originally began through efforts of SIL personnel resident in those communities. NPMTTA, in collaboration with SIL, Translators Association of the Philippines (TAP), and the Philippine Bible Society, sponsors series of workshops to provide their personnel with new skills and update existing skills related to translation and language development. NPMTTA is also working with TAP and SIL to sponsor a number of scholars for graduate level studies in the Applied Linguistics Program at Alliance Graduate School in Manila, which offers a nationally accredited Graduate Diploma in Applied Linguistics, as well as a Master of Arts in Applied Linguistics with specializations in either Language Development or Bible Translation.

In addition to providing relevant training for members of local language communities, SIL is committed to assisting members of those communities to gain a voice in a broader context. SIL does this by co-hosting a variety of conferences and workshops with other agencies. In 2006, SIL and the Linguistic Society of the Philippines hosted the Tenth International Conference on Austronesian Linguistics.<sup>36</sup> Because the conference was held in the Philippines, a record number of Filipino scholars were able to participate and present their research findings on Philippine languages. A special session on languages of Palawan was sponsored at that conference by faculty of the Palawan State University. In 2003, SIL, Mahidol University and UNESCO sponsored a landmark conference on Language Devel-

---

<sup>36</sup> All papers submitted by participants of 10-ICAL are available on-line at <http://www.sil.org/asia/philippines/ical/papers.html>. Edited collections of selected papers are to appear in the 2006 volumes of *Philippine Journal of Linguistics*, and in several volumes of the series *Studies in Philippine Languages and Cultures*.

opment, Language Revitalization, and Multilingual Education in Minority Communities in Asia.<sup>37</sup> Follow-up workshops are equipping policy makers, administrators and classroom teachers to bring local languages into the classroom as media of instruction.

In May 2007, SIL International – Asia signed a memorandum of understanding with UNESCO Bangkok – Asia and Pacific Regional Bureau for Education which included joint commitments to the following efforts of particular relevance to endangered Austronesian languages:

- building the capacity of educational policy makers to make decisions with regard to language-in-education issues.
- advocating for the affirmation and support of linguistic diversity and the recognition of the danger that many existing languages will be lost.
- exploring further areas of cooperation such as publications in endangered languages, software for universal access to information and knowledge, and other areas of potential common interest.

The last point of this MOU mentions a crucial area for the documentation and revitalization of endangered languages not yet considered in this paper—that of software for language development.

**8.3 SOFTWARE.** SIL International has developed over sixty pieces of software to support the work of language development, most of which are available for free download.<sup>38</sup> Programs available to record and analyze survey data, teach the International Phonetic Alphabet, parse words and interlinearize texts, adapt text material from one language into another, build dictionaries, use special fonts and characters, and more. *The Linguist's Shoebox* is an integrated data management and analysis tool especially helpful for building a dictionary based on interlinearized text. *The Field Linguist's Toolbox*, built on *Shoebox 5*, is an upgrade of *Shoebox* which adds Unicode support.<sup>39</sup> *Discourse Profiler*, developed by Phil Quick (discussed in his paper in this volume), builds on *Toolbox* and is designed for analyzing the context of specific discourse features. SIL has also produced an electronic library of many of the linguistic resources commonly used in SIL fieldwork.<sup>40</sup>

*SIL Fieldworks* is the most recent suite of language software developed by SIL for the management of language and cultural data. *Fieldworks* assists the field linguist with tasks associated with the initial collection of data, with the analysis of linguistic rules and cultural precepts, and with the publication of findings. It facilitates categorizing cultural

---

<sup>37</sup> The conference program, papers and other information are available at <http://www.sil.org/asia/ldc/>.

<sup>38</sup> See <http://www.sil.org/computing/catalog/index.asp> for SIL's on-line software catalog and instructions for downloading or ordering.

<sup>39</sup> For further information, see also <http://www.discourseprofiler.com>.

<sup>40</sup> *LinguaLinks Library*, which contains 149 book or book-length works, and over 200 SIL journals, is available for purchase at [http://www.ethnologue.com/LL\\_docs/contents.asp](http://www.ethnologue.com/LL_docs/contents.asp).

observations based on the Outline of Cultural Materials and incorporates a Dictionary Development Process for building and refining a dictionary. *Fieldworks* is a sophisticated software tool to assist trained linguists and other researchers of language and culture.

SIL is also developing tools designed for non-linguist members of local communities to assist directly in language development projects through data gathering. An example of this sort of tool is *WeSay*, an Open Source project being developed by the Software Unit of the Payap University Linguistics Institute and SIL International (discussed in this volume by Eric Albright and John Hatton).<sup>41</sup> *WeSay* is envisioned to be an extremely simple, task-oriented tool that will run on inexpensive rugged computers suitable for use in a rural environment. *WeSay* is especially intended for building dictionaries and text collections, in collaboration with advisors who can handle the more technical aspects of the analysis and publication of data. Straightforward import and export facility between *WeSay* and *Fieldworks* is planned.

SIL supports linguists and language communities through the development of affordable, appropriate language development technology. Worthy of special mention here is another technology for producing reading materials that has been used successfully by local language communities in various parts of the world. *Shellbook Publishing Systems*<sup>42</sup> provides software that enables local users to insert culturally appropriate text and illustrations into a template for an existing publication. Through this method, high quality materials with accurate vital information can be produced as a supplement to locally created literature.

**9. CONCLUDING REMARKS.** This paper has briefly reviewed the history of SIL International language development work in Austronesian languages, giving special consideration to issues of language endangerment and revitalization. Using the Malaysian Indigenous Peoples Conference on Education as a microcosm of today's Austronesian language communities, it shows that these communities are generally multilingual, increasingly linked to each other and to the outside world, ready to learn from the experiences of others, and often ready to commit to promoting and developing their own languages. Numerous factors influence the degree of vitality or endangerment of a particular language. Many researchers hold that literacy in a local language is a crucial factor for a language's survival in today's world. The most critical issue, however, is whether parents are passing on the language of the community to children in the home. Assuming that parents are still passing on their heritage language to their children, many of the language development products and activities that SIL has traditionally helped produce can be considered as prerequisites for language revitalization, or in the words of Batibo (2007), "language empowerment". Finally, this paper offered some ways that SIL may best assist those working on behalf of endangered Austronesian languages today—namely, through efforts in documenting and describing, consulting and training, and providing affordable, appropriate language software.

---

<sup>41</sup> For further information, see <http://www.wesay.org>.

<sup>42</sup> *Shellbook Publishing Systems* is a private corporation, not part of SIL. For further explanation of the *Shellbook* philosophy and methodology, see <http://www.shellbook.com/about.faces>.



The very fact that computer software enters into a discussion on language revitalization indicates that we are in a different world in this twenty-first century than when SIL began language development work in Austronesian languages in the 1950s. At the same time that national and international languages and the forces of globalization are threatening to wipe out the majority of the world's local languages, technology also offers some means for strengthening the role and status of endangered languages. Still, in the words of Joshua Fishman, when it comes to promoting the vitality of local languages, the most important "laptop" is the lap of the parent or grandparent in passing on one's intangible heritage to the next generation.<sup>43</sup>

Like today's local language communities, SIL is ready to learn from and link with others as we serve the parents and grandparents, children and grandchildren who desire to preserve and promote their increasingly endangered languages.

---

<sup>43</sup> Fishman 1996: "My wife engages in laptop publishing. She publishes in the Yiddish language for our grandchildren. But let me tell you, the true lap top here is my lap and her lap and the laps of the children's mother and father. That is a bond with the language that will stay with them after we are long gone. That is the lap top of language."

## REFERENCES

- AUSTIN, PETER K. 2006. Data and language documentation. In *Essentials of language documentation*, ed. by Gippert, Himmelmann and Mosel, 87-112. Berlin: Mouton de Gruyter.
- AWID, MELINDA T. 2001. Literacy for development: A Western Subanon experience. E-publication of the International Literacy Institute, July/ August 2001. <http://www.literacy.org/li/li200108/awid.html>
- BAKER, COLIN. 2006. *Foundations of Bilingual Education & Bilingualism*, 4th edition. Clevedon: Multilingual Matters.
- BATIBO, HERMAN M. 2007. Language documentation as a strategy for the empowerment of the minority language of Africa: Playing the devil's advocate. Paper delivered at the joint meeting of the African Language Teachers Association and Annual Conference on African Linguistics, March 21-25, 2007, University of Florida at Gainesville.
- BLOOMFIELD, LEONARD. 1917. *Tagalog texts with grammatical analysis*. Urbana, IL: University of Illinois.
- BREZZINGER, MATTHIAS, ET AL. 2003. Language vitality and endangerment. Paris: UNESCO Expert Meeting on Safeguarding Endangered Languages. [http://portal.unesco.org/culture/en/file\\_download.php/947ee963052abf0293b22e0bfba319cclanguagevitalityendangerment.pdf](http://portal.unesco.org/culture/en/file_download.php/947ee963052abf0293b22e0bfba319cclanguagevitalityendangerment.pdf)
- CAHILL, MICHAEL. 1999. From endangered to less endangered: Case histories from Brazil and Papua New Guinea. SIL Electronic Working Papers. <http://www.sil.org/silewp/2004/silewp2004-004.htm>
- CANTONI, GINA, ed. 1996. *Stabilizing indigenous languages*. Flagstaff: Northern Arizona University. <http://www.ncela.gwu.edu/pubs/stabilize/index.htm>
- DEKKER, DIANE E. 1999. Leaving home - my language or yours. In *The Filipino Bilingual: A multidisciplinary perspective*, ed. by Maria Lourdes S. Bautista and Grace O. Tan, 103-111. Manila: Linguistic Society of the Philippines.
- DEKKER, DIANE E. and ROSE C. DUMATOG. 2003. First language education in Lubuagan, Kalinga, northern Philippines. Paper presented at the Conference on Language Development, Language Revitalization and Multilingual Education in Minority Communities in Asia. Bangkok, Thailand. November 6-8, 2003. [http://www.sil.org/asia/ldc/parallel\\_papers/dumatog\\_and\\_dekker.pdf](http://www.sil.org/asia/ldc/parallel_papers/dumatog_and_dekker.pdf)
- DEKKER, DIANE E. and Catherine Young. 2005. Bridging the gap: The development of appropriate educational strategies for minority language communities in the Philippines. *Current Issues in Language Planning* 6(3):182-199.
- FISHMAN, JOSHUA A. 2000. Reversing language shift: RLS theory and practice revisited. In *Assessing ethnolinguistic vitality: Theory and practice: Selected papers from the Third International language Assessment Conference*, ed. by Gloria Kindell and M. Paul Lewis, 1-25. Dallas: SIL International.
- FISHMAN, JOSHUA A. 1996. What do you lose when you lose your language? In *Stabilizing indigenous languages*, ed. by Gina Cantoni. Northern Arizona University. <http://www.ncela.gwu.edu/pubs/stabilize/iii-families/lose.htm>
- FISHMAN, JOSHUA A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, UK: Multilingual Matters.
- GALLMAN, ANDREW F. 1997. *Proto East Mindanao and its internal relationships*. LSP Special Monograph 44. Manila: Linguistic Society of the Philippines.
- GIPPERT, JOST, NIKOLAUS P. HIMMELMANN and ULRIKE MOSEL, eds. 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.

- GRIMES, BARBARA F. 2001. Global language viability. In *Endangered languages of the Pacific rim: Lectures on endangered languages 2*, ed. by Osamu Sakiyama, 45-68. ELPR Publication Series C002. Osaka, Japan: ELPR. <http://www.sil.org/SOCIOLX/ndg-lg-grimes.html>
- GRIMES, JOSEPH E. 1995. Language endangerment in the Pacific. *Oceanic Linguistics* 34: 1-12.
- GRIN, FRANÇOIS. 2005. Linguistic human rights as a source of policy guidelines: A critical assessment. *Journal of Sociolinguistics* 9(3): 448-460.
- HANS RAUSING ENDANGERED LANGUAGES PROJECT. 2004. Language documentation: What is it? <http://www.hrelp.org/documentation/whatisit/>
- HEADLAND, THOMAS N. 2003. Thirty endangered languages in the Philippines. *Work Papers of the Summer Institute of Linguistics*, University of North Dakota Session 47. <http://www.und.edu/dept/linguistics/wp/2003Headland.PDF>
- HIMMELMANN, NIKOLAUS P. 2006. Language documentation: What is it and what is it good for? In *Essentials of language documentation*, ed. by Gippert, Himmelmann and Mosel, 1-30. Berlin: Mouton de Gruyter.
- HIMMELMANN, NIKOLAUS P. 2002. Documentary and descriptive linguistics (full version). <http://www.hrelp.org/events/workshops/eldp2005/reading/himmelmann.pdf>
- HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36.161-195.
- HIMMELMANN, NIKOLAUS P., Jost Gippert and Ulrike Mosel. 2006. Editor's preface. In *Essentials of language documentation*, ed. by Gippert, Himmelmann and Mosel, v-vii. Berlin: Mouton de Gruyter.
- KINDELL, GLORIA and M. PAUL LEWIS, eds. 2000. *Assessing ethnolinguistic vitality: Theory and practice: Selected papers from the Third International Language Assessment Conference*. Dallas: SIL International.
- KLAUS, DAVID. 2003. The use of indigenous languages in early basic education in Papua New Guinea: A model for elsewhere? *Language and education* 17(2): 105-111.
- LANDWEER, MARTHA LYNN. 2007. Indicators of ethnolinguistic vitality. <http://www.sil.org/SOCIOLX/ndg-lg-indicators.html>
- LANDWEER, MARTHA LYNN. 2006. *A Melanesian perspective on mechanisms of language maintenance and shift: Case studies from Papua New Guinea*. Ph.D. thesis, University of Essex.
- LEWIS, M. PAUL. 2006. Towards a categorization of endangerment of the world's languages. SIL Electronic Working Papers, SILEWP 2006-002. <http://www.sil.org/silewp/abstract.asp?ref=2006-002>
- LEWIS, M. PAUL and BARBARA TRUDELL. In press. Language cultivation in contexts of multiple community languages. In *Handbook of educational linguistics*, ed. by Bernard Spolsky and Francis M. Hult. Oxford: Blackwell.
- LITTERAL, ROBERT. 1999. Language development in Papua New Guinea. SIL Electronic Working Papers, SILEWP 1999-002. <http://www.sil.org/silewp/1999/002/silewp1999-002.html>
- LITTERAL, ROBERT. 2004. Vernacular education in Papua New Guinea. Background paper prepared for the Education for All Global Monitoring Report, 2005. <http://unesdoc.unesco.org/images/0014/001466/146660e.pdf>
- McFARLAND, CURTIS D. 1981. *A linguistic atlas of the Philippines*. Monograph series, 15. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa.

- MILLER, CAROLYN P. 2000. Modifying language beliefs: A role for Mother-Tongue advocates? In *Assessing ethno-linguistic vitality: Theory and practice*, ed. by Gloria Kindell and M. Paul Lewis, 167-187. Dallas: SIL International.
- PIKE, KENNETH L. 1967. *Language in relation to a unified theory of the structure of human behavior*. *Janua Linguarum, series maior*, 24. The Hague: Mouton.
- PIKE, KENNETH L. 1982. *Linguistic concepts: An introduction to tagmemics*. Lincoln: University of Nebraska Press.
- PITTMAN, RICHARD S. 1998. Back from the brink: Sixteen accounts of ethnic renaissance. Waxhaw, NC: Summer Institute of Linguistics.
- PROGRAMA LENGUA RAPA NUI. ms. Resultados del programa lengua Rapa Nui, 1977-2006. Pontificia Universidad Católica del Valparaíso and SIL International.
- ROSS, MALCOLM D. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Canberra: Pacific Linguistics. C-98.
- Scebold, Robert N. 2003. *Central Tagbanwa: A language on the brink of extinction: Sociolinguistics, grammar, and lexicon*. Manila: Linguistic Society of the Philippines.
- SPOLSKY, BERNARD and FRANCIS M. HULT, eds. In press. *The handbook of educational linguistics*. Oxford: Blackwell.
- TOWNSEND, WILLIAM CAMERON. 1952. *Lázaro Cárdenas: Mexican democrat*. Ann Arbor: George Wahr.
- WALTON, CHARLES. 1979. A Philippine language tree. *Anthropological Linguistics* 21(2): 70-98.
- WEBER, NANCY THIESEN DE and ROBERTO WEBER CH. 1998. Diagnóstico lingüístico de los educandos de kinder a séptimo año básico de la Escuela Lorenzo Baeza Vega de Isla Pascua al final del año escolar 1997: Informe y recomendaciones.
- WEBER, DIANA DAHLIN, DIANE WROGE and JOAN BOMBERGER YODER. 2007. Writers' workshops: A strategy for developing indigenous writers. *Language Documentation & Conservation* 1(1):77-93. <http://nflrc.hawaii.edu/ldc/>
- WRIGGLESWORTH, HAZEL J. In press. *Narrative episodes from the Tulalang epic*. Manila: Linguistic Society of the Philippines.
- WRIGGLESWORTH, HAZEL J. and AMPATUAN AMPALID. 2004. *The song from the mango tree*. Manila: Linguistic Society of the Philippines.
- Wrigglesworth, Hazel J. and PENGENDÀ MENGENGILID. 1993. *Good character and bad character/The Manobo storytelling audience as society's jurors*. Manila: Linguistic Society of the Philippines.
- WRIGGLESWORTH, HAZEL J. 1991. *The maiden of many nations: The skymaiden who married a man from earth*. Manila: Linguistic Society of the Philippines.
- WROGE, DIANE. 2002. Papua New Guinea's vernacular language pre-school programme. UNESCO Policy Brief on Early Childhood, 7. <http://unesdoc.unesco.org/images/0013/001373/137383e.pdf>
- YOUNG, CATHERINE. 2002. First language first: Literacy education for the future in a multilingual Philippine Society. *International journal of bilingual education and bilingualism* 5(4): 221-232.
- ZORC, R.DAVID. 1977. *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Pacific Linguistics, C-44. Canberra: Australian National University.

J. Stephen Quakenbush  
 steve\_quakenbush@sil.org

# 5

## Local Autonomy, Local Capacity Building and Support for Minority Languages: Field Experiences from Indonesia

I Wayan Arka

*The Australian National University*

This paper discusses the complexity of language/cultural maintenance and revival, highlighting the significance of building and supporting long-term local capacity. These complex issues are discussed in the current context of rapid political change towards greater local autonomy in Indonesia. After some background on aims and regulations of decentralization, the Balinese in Bali and Rongga in Flores are compared and discussed based on the author's field experiences. It is argued that capacity building and support must include more than simply developing human resources. Strengthening, reforming, and/or restoring relevant institutions, particularly in relation to customary *adat* systems, are equally important. While a macro perspective must be adopted, priority must be given to a community-based approach and to long term capacity building and support at the most local level. The comparison of the Rongga and Balinese helps clarify how a range of inter-related socio-political and economic variables at the local and regional levels play a significant role in providing and/or inducing good conditions for bottom-up community-based initiatives in language/cultural maintenance and revival.

**1. INTRODUCTION.** The need for capacity building in maintaining and reviving indigenous cultures/languages has been highlighted recently (de Graff and Shiraishi 2004; Dimmendaal 2004; Foley 2004; Lastra 2004, among others). In this paper, I discuss issues of capacity building that might affect the wellbeing of local cultures/languages in Indonesia. This is done in relation to the rapid political change (decentralization) currently being experienced. I address the aims, regulations and current implementation of the current drastic decentralization laws. The description is based on field experiences from my current language documentation project of Rongga in Manggarai Flores Indonesia.<sup>1</sup>

In order to give a more detailed appraisal, it is necessary to look at what has happened elsewhere in Indonesia. Comparing Rongga in Manggarai with other local languages across Indonesia is too big an undertaking for the present paper, and in this paper I focus

---

<sup>1</sup> Rongga Documentation Project (2004-2006) was funded by a grant from the Rausing Fund SOAS London awarded to the author (IPF 0011). I thank Rongga people, in particular Juventus Rau, Franciscus Seda, Bapak Anton Gelang, and Yohanes Nani, for their generosity and hospitality during my fieldwork in Flores. I also thank the audience at the International Conference on Austronesian Endangered Language Documentation, Taichung Taiwan (June 2007) for their valuable comments on the earlier version of the paper. Special thanks go to two anonymous reviewers for their insightful comments and questions, and Victoria Rau for her time and patience in reading and checking the final version of the paper thoroughly. All errors are, however, mine.

on the comparison of Rongga in Manggarai, Flores Island, NTT (*Nusa Tenggara Timur*, or East Nusa Tenggara) province with Balinese in Bali. The specific reasons for choosing Balinese are elaborated on in subsection 5.2.

I wish to highlight the following points. Firstly, capacity support is as important as capacity building. Secondly, capacity related variables that support viable indigenous cultures/languages are complex, and capacity building and support must include more than simply developing human resources. It must also include strengthening, reforming, or restoring relevant institutions/organizations. In the Indonesian context, this also means reviving and strengthening the traditional customary *adat* system.

Thirdly, while capacity building and support must be approached from a broad or macro, perspective, with a top-down and bottom up approach at the same time, I argue for the significance of priority given to the capacity building and long term support at the regional/local level, particularly at the village level. It will be shown that the current climate of democracy in Indonesia with the introduction of new laws on regional/local autonomy has provided new promising prospects of revival of local cultures/languages. While this paper is drawn from Rongga and Bali experiences, I believe that the insights and recommendations formulated can be extended to minority cultures/languages in other parts of Indonesia.

This paper is organized in the following way. To provide the background, I present basic facts about Indonesia's geography, demography, government and political systems in section 2, followed by a brief historical context of the introduction of the new autonomy laws in section 3. After discussing the goals of local autonomy (section 4) and the significant changes in the new law (section 5.1), I discuss the implications of the laws in relation to the well-being of indigenous cultures and languages on the basis of field experiences in Balinese and Rongga cases (section 5.2). Section 6 provides the summary and final remarks.

## 2. BASIC FACTS ABOUT INDONESIA IN BRIEF

**2.1 GEOGRAPHY AND SOCIETY.** Indonesia is the largest archipelagic nation in the world with a total of 17,508 islands spreading between Asia and Australia. Among these, about 6,000 have names and around 1,000 are permanently settled. It has the world's fourth largest population after China, India, and the United States, totaling an estimated 225.3 million people in 2005 (UN, 2005).

Ethnically, culturally, and linguistically, Indonesia is highly diverse. There are an estimated over 350 ethnolinguistic groups in Indonesia. There are 731 local (Austronesian and non-Austronesian) languages in Indonesia (SIL International 2001), roughly one-tenth of all the languages in the world today. Some have large numbers of speakers, e.g. Javanese (75 million), Sundanese (27 million) and Madurese (nearly 14 million) (see, Steinhauer 1994; Sneddon 2003:198). There are other smaller languages with over 1 million speakers, e.g. Minangkabau (6 million), Buginese (3.6 million), Balinese (3.2 million), and Acehese (2.4 million). However, there are also many small languages with speakers in the thousands, e.g. Rongga (5000 speakers).

The cultural and ethnic diversity has been in the past, and still will be, a challenge for any government to maintain a united Indonesia that is politically stable and economically prosperous across the archipelago. In addition, tension due to religious issues and terrorism further complicates the situation.

**2.2. GOVERNMENT AND ADMINISTRATIVE UNITS.** 1 shows the structural and territorial government of Indonesia within the new framework of regional/local autonomy. Indonesia is divided into autonomous provinces, which consist of districts or regencies (*kabupaten*) and city municipalities (*kota madya* or *kodya*). Districts and municipalities are technically the same level of government but they are distinguished by the location of government administration: *kota* (municipality) in urban area vs. *kabupaten* (district) in rural areas. Within districts and municipalities there are sub-districts (*kecamatan*), which are smaller administrative government units. Each sub-district is further divided into villages.

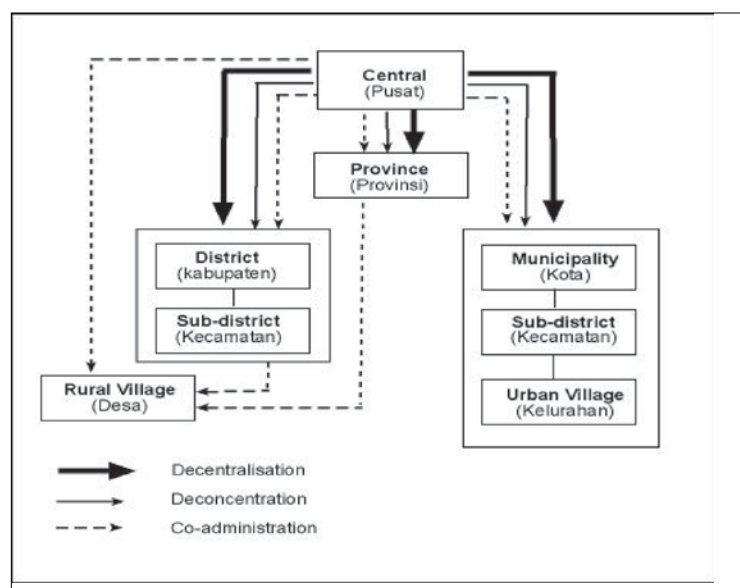


FIGURE 1. Structural government and administration of Indonesia according to the autonomy law 22/1999<sup>2</sup> (adapted from Usman (2002))

Villages in rural areas are called *desa*, while in urban areas they are called *kelurahan*. In Bali, however, there are two types of village units which may co-exist in the same areas: the government administrative unit of *desa dinas* and the traditional village unit of *desa*

<sup>2</sup> The new law for local governments (Law 32/2004) was introduced in 2004 to replace Law 22/1999 in response to the amendments to the 1945 constitution, especially with regards to direct elections and related issues emerging in 1999-2004. The structural hierarchy of administrative units with respect to local autonomy at the district level depicted in Figure 1 basically remains the same. Law 32/2004, however, appears to place the rural village (*desa*) in the same structural dominance as the

*adat*. The retention of *desa adat* has been critically important for the maintenance of Balinese culture/language even though the dualism has sometimes resulted in rivalry between the two (see, Warren 1993).

Currently there are 33 provinces, 370 districts and municipalities, 5,263 sub-districts, 7,113 *kelurahan*, and 62,806 villages in Indonesia. These figures will certainly increase because there has recently been a growing tendency of *pemekaran* (formation of new provinces and districts).

The Indonesian constitution of 1945 regulates separation of powers among the executive, legislative, and judicial branches. The president holds executive power. The president is both chief of state and head of government. The 1945 constitution was amended in August 2002. One important amendment is a direct presidential election, beginning with the 2004 general election. Prior to the 2004 elections, the People's Consultative Assembly (*Majelis Permusyawaratan Rakyat*—MPR) chose the president and vice president. Under the revised election laws, governors (*gubernur*), mayors (*walikota*), and district heads (*bupati*) are also now directly elected. The *kepala desa* (rural village head) has been traditionally directly elected by the village community. However, the sub-district head (*camat*) and the urban village head (*lurah*) are civil servants appointed by the local district government.

At the regional/local level, the chief executives are provincial governors, district heads, mayors of cities, and village heads. Regional/local legislation is handled by provincial and district parliaments, called DPRD (*Dewan Perwakilan Rakyat Daerah*)

**2.3. NATIONAL AND LOCAL POLITICS.** Indonesia has enjoyed freedom and democracy in the post-Soeharto era, starting in 1998. There is now freedom of speech, including freedom of the press. There is no longer controlled political development as in the *Orde Baru* (New Order) era.<sup>3</sup> As a result, there has been an explosion in the number of political parties. While there are numerous parties vying for power, none enjoys national majority support. The main parties include the Golkar (Functional Group), Party Crescent Moon and Star Party (PBB), Democratic Party (PD), Indonesia Democratic Party of Struggle (PDI-P), National Awakening Party (PKB), National Mandate Party (PAN), Prosperous Justice Party (PKS), and United Development Party (PPP).

At the local/regional levels, especially in non-Muslim electorates such as Bali and Manggarai/NTT, Golkar and PDI-P are generally the two most popular parties. In recent developments, an independent candidate was allowed in Aceh and won the election for governor (Aceh is a special case, and an independent candidate is not (yet) permitted in other local electorates in Indonesia).

*kelurahan*. This change is a drawback to the spirit of traditional autonomy of *desa* (see 0).

<sup>3</sup> The *Orde Baru* (New Order) era refers to the period of 1966-1998 when Soeharto was in power in Indonesia. This term was used by Soeharto to contrast his regime with that of his predecessor, Soekarno. However, it has come to be used pejoratively since it is associated with corruption, collusion and nepotism in Soeharto's authoritarian regime.



Money, politics, ethnicity and/or clan base, and religious affiliation often play an important role in national and local politics. There are mass campaign rallies as a show of force in the lead up to elections. Speeches by the candidates generally have no clear or fresh proposals that address national or local issues.

**3. HISTORICAL CONTEXT OF THE NEW AUTONOMY LAWS.** Indonesia's political system was highly centralized under the Soeharto regime. Since independence from the Dutch in 1945, the central government has struggled for decades to contain separatist rebellions. Soeharto, an army general, believed in, and therefore implemented, a highly centralized powerful system to maintain the unity of Indonesia, by force if necessary. During his *Orde Baru* (New Order) era, any measure was regarded as legitimate in the name of the Indonesian nation and development (*pembangunan*).

The negative impact of Indonesianisation on minority cultures/languages under the centralized and autocratic system in the Soeharto era has been well documented. Soeharto's style of governance has also produced a highly corrupt system, which was ultimately the cause of his downfall in 1998. The Indonesian economy collapsed following the economic crisis spreading in East and South-East Asia the preceding year. Soeharto's downfall closed his New Order era, and marked the beginning of the reformation era.

Since the fall of Soeharto in 1998, four Presidents have been elected: B.J. Habibie (1998-1999), Abdurachman Wahid (1999-2001), Megawati Sukarno Putri (2001-2004), and Susilo Bambang Yudoyono (2004-present). In response to public demand and the spirit of reformation, Indonesia has embarked on a decentralization program on a grand scale. Two laws adopted by the Parliament in early 1999 required the government to implement drastic measures for decentralization within two years. The Law of Regional Governance (Law 22, 1999) regulates political and administrative responsibilities for the local government. The Law of Fiscal Balance (Law 25, 1999) delineates, among others, new distribution of revenue sources including sharing of oil and gas revenues between central and local governments. The new policy of regional autonomy became effective on January 1<sup>st</sup>, 2001.<sup>4</sup> There have been significant changes since then, especially the amendments of the Indonesian constitution of 1945 and the introduction of new laws related to political parties (31/2002), 12/2003 and on (direct) elections, etc. In response to these swift changes, law 32/2004 on local government was introduced as a revision of law 22/1999. The relevance and impact of this law is discussed in some detail in sections 4 and 5 below.

**4. WHY LOCAL AUTONOMY FOR INDONESIA?** There are historical, geographical, and socio-political reasons for the implementation of local autonomy in Indonesia (cf., Said 2005:78-88). Geographically, as described in §2, Indonesia consists of thousands of islands, and hundreds of ethnic groups with different languages and cultures. The population of over 220 million people is spreading in vast areas and often on remote islands which demands varying approaches and services in response to quite diverse needs. A

---

<sup>4</sup> Decentralisation was in fact already initiated when Soeharto was still in power, e.g. decentralisation of education with the introduction of *muatan local* (*mulok*), or the Local Content Curriculum, which started in 1994 with discussions about it already taking place in the 1980s (Bjork 2004). This was encouraged by the International donors. It should be noted that decentralisation during Soeharto's time was not of the nature and scale proposed in the reformation era.

decentralized system of governance is naturally a logical system for such a geographic and demographic situation. It is hard to imagine how a centralized system could respond and offer satisfying services to the range of local needs across Indonesia.

Historically, the founding fathers of Indonesia, when debating the 1945 constitution and considering the country's geography and demographic diversity, came to the conclusion that a decentralized system was ideal for an independent Indonesia. Hence, article 18 of the 1945 constitution was formulated to recognize local diversity and guarantee that local needs would be well taken care of.

While local autonomy was already conceived, it gained momentum in 1999 when Soeharto was suddenly forced out of power, after the collapse of the national economy. The euphoria of reformation led to the reassessment of Soeharto, in particular his centralized system of government during his 30 years in power. Many believed that much of the problem was rooted in his centralized and dictatorial style of government. Hence, the new Law on Regional Governance (Law 22, 1999) was hastily passed (formally authorized on 4 May, 1999) with the high expectations that the law could fix the problems. The preamble (i.e. the consideration section) of the new law explicitly states:

- That the 1945 constitution guarantees local autonomy, i.e. the freedom of the local government to manage its local affairs;
- That local autonomy needs to be implemented to face new challenges nationally or internationally, with the principles of democracy, community participation and empowerment, equity and justice, recognition of diversity within regions;
- That the existing legislatures, namely laws no. 5 1974 on *Pokok-pokok Pemerintahan di Daerah* and Laws no. 5 1979 on *Pemerintahan Desa*, are outdated. The latter in particular imposed a uniformed local system in name and structure at the village level across Indonesia, against the 1945 constitution, and therefore had to be replaced.

Politically, the central government expects that local autonomy can address the long-term and deep regional discontents against the system that is (perceived as) too centralized by Jakarta. The political separatist agenda of those disgruntled regions (mainly outside Java) is expected to be accommodated and the breakup of Indonesia can hopefully be avoided.

In short, there are three aims of the new law of local autonomy. The first one is to correct and replace the past regulations for regional/local governments. The second aim is to bring about better delivery of government services to meet local needs, and to raise the level of local government accountability within the principles of democracy, community participation and empowerment, equity and justice, recognition of diversity and potential within regions. Finally, related to these two aims, local autonomy will address political discontent and maintain the unity of Indonesia as a nation.

**5. WHAT DOES LOCAL AUTONOMY MEAN FOR THE LANGUAGE / CULTURAL MAINTENANCE?** Theoretically, local autonomy would open an opportunity for the local government to take care of its own interests. In practice, however, the implementation

of local autonomy has not been as smooth as expected. The slow progress is due to a range of local and national problems, such as mediocre economic growth and other political concerns of separatism, terrorism and uncertainties of clear future directions (Turner *et al.* 2003:145). The readiness of the local governments to exercise new responsibilities with greater autonomy is also a factor. In addition, decentralization has not been implemented whole-heartedly due to inherent conflicts between the goals of decentralization policies and the interests of competing groups or factions within the (central) government (who then often actively resist the implementation of new autonomy measures) (see Bjork 2004).

Relevant for our discussion is the implication of the laws in relation to the well being of the indigenous communities, especially the maintenance (and possible revival) of indigenous cultures and languages. This prospect varies across Indonesia, depending on a range of factors at the regional/local level. This is discussed in relation to Balinese and Rongga in subsection 5.2, but significant changes in the autonomy law are first given in 5.1 below.

**5.1 SIGNIFICANT CHANGES IN THE AUTONOMY LAWS.** Law 22/1999 contains 134 articles, and is intended to replace law number 5/1974 on local government. Included in 22/1999 are 19 articles as amendments of law 5/1979 regarding village governance (*Pemerintahan Desa*). In addition to this law, another law (Law 25/1999), which contains 33 articles, regulates the fiscal balance between the central and local governments. Law 32/2004 is the revised version of laws 22/1999 and 25/1999 in response to changes and related issues emerging in 1999-2004.

The first major change in the new laws is the change in the regional hierarchy. Districts (*kabupaten*) and cities (*kota*) hold equivalent autonomous status. Unlike in the previous system where they are hierarchically organized under the province (*propinsi*), districts and cities are an independent focus of regional governments with direct communication to central agencies in Jakarta. (The position of province is rather unclear in the current hierarchy, which creates problems in coordinating the district governments within their jurisdiction).

The second change is the acquisition of quite a wide range of autonomous functions awarded to the regional governments at the district and city level, but not with provinces. The functions and responsibilities awarded include public work, health, education and culture, agriculture, communication, industry and trade, capital investment, environment, land, cooperatives, and manpower affairs and the management of national resources. (Note that education and culture (with local language included) are now in the hands of the local government). The management of the wide range of functions needs to be backed up with good (local) financial support, human resources and infrastructure. Unfortunately, the necessary backup is not there (yet). We will return to the issues relevant for language/cultural maintenance below.

It should be noted that local autonomy laws also regulate the structure, autonomy, and democratic governance at the rural village (*desa*) level. The village officials who manage various aspects of rural governance, economic development, and finances consist of a village chief, a secretary, and the board of village representatives. The rural village head is elected directly. The board of village representatives is called BPD or *Badan Perwakilan Desa* (Village Representative Board) according to Law 22/1999. Its members are directly elected by the community, and the village chief must be accountable to the community via the BPD.

In 2004, however, this was changed in Law 32/2004: BPD stands for *Badan Permusyawaratan Desa* (Village Consultative Board); its members are appointed; and the village chief must be held accountable to the district head. This could be regarded as a setback, and against the spirit of democracy and local autonomy at the village level because it would mean that *desa* is the extension of the government and thus could be politicized and controlled by the district government just as in Soeharto's era. In order to assess the implementation and implication of the local autonomy, I now turn to Balinese in Bali and Rongga in Flores.

**5.2 BALINESE VS. RONGGA.** The Balinese language and culture is not in any way threatened or endangered. Balinese is a relatively large language, spoken by some three million people. It is used in a range of domains in daily communication in Bali, except in formal educational settings. (In this case, Indonesian is the language of instruction.) Balinese has had a literacy tradition for over a millennium (Arka 2006), and enjoys strong support from the Balinese community and local government in Bali, further discussed in 5.2.1 below.

In contrast, Rongga is a marginalized minority language (with around 5,000 speakers). It is increasingly under pressure not only from Indonesian, but also from the local regional language of Manggarai and its stronger neighbor Waerana. While Rongga is still relatively healthy in the interior part of its territory, e.g. in the hamlet of Nangarawa, it is increasingly threatened in certain areas in the northern part of Rongga areas along the Trans Flores road. Language shift to Waerana has been reported in these areas. Further discussions on the sociolinguistic situations of Rongga in Manggarai is given in Arka (2005).

Balinese is chosen in this paper for the comparative discussion of local autonomy and its implication for local cultures/languages for the following reasons. Firstly, Bali can be regarded as a model of a success story of local autonomy that has further enhanced local capacity for culture and language maintenance/revival. Secondly, Balinese contexts illustrate the complexity of supporting variables at the local level, in particular the significance of retaining the *adat* (customary) system at the village level. Third, I am a Balinese and therefore in a unique position for this study: I am familiar with the situation in Bali, and have also been doing fieldwork research on Rongga in Flores. Much can be learned from the Balinese case, and this comparison is expected to lead to a better understanding of the issues involved and the nature of challenges ahead in the context of new decentralized modern Indonesia.<sup>5</sup>

**5.2.1 BALINESE IN BALI.** Balinese enjoys strong support from the local communities, local government and local NGOs, and is backed up by good human and financial resources. Traditional social structures and traditional leadership are still mostly intact, providing good conditions for maintenance of tradition where art and language are a significant part of the Hindu-based cultural practices of the Balinese.

---

<sup>5</sup> Ideally a comparison would be done with a minority language, having similar conditions as Rongga which has benefited and done well under the new autonomy laws. Such a minority language is to my knowledge not (yet) extant in Indonesia.

The reasons why Balinese enjoys strong support from the provincial and district governments are that it is spoken in all districts in the province of Bali, and that the governments at these levels are controlled by the Balinese. Hence, programs to promote Balinese can be well coordinated at all levels, from the provincial level to the district, sub-district, and even down to village levels.

A deliberate government program of reviving Balinese culture and language was in fact started in the 1980s when a Balinese anthropologist, Ida Bagus Mantra, became the governor of Bali. One of his programs which continue today is the annual Bali Art and Culture Festivals, generally for one month from June to July. In these festivals, there are exhibitions, workshops/seminars, competitions, and performances of different kinds of arts and cultural practices across Bali, including *lontar* (palm-leaf) reading or *makakawin/makabebasan*.

The festivals are intended to highlight the richness of Balinese culture, and to revive certain endangered cultural practices. Hence, these are occasions where rare or unique traditions from different parts of Bali are exhibited or performed. While these festivals are not mainly organized for tourist purposes, they are also good tourist attractions, generating income for the local government.

In the current local autonomy reformation era, support from the provincial and district governments for Balinese culture is continuing, and even improving. The districts of Badung and Gianyar are two rich districts in Indonesia. The revenues mainly come from tourism. The awareness of democracy (i.e. the local government to be elected by the local people to serve local needs) has led to a change towards a better distribution of government funding and services addressing local needs. In particular, there has been growing awareness of sustainable cultural tourism in Bali: in order to continuously benefit from tourism (and tourism-related industries), unique local cultures must be maintained and revived because these are the main tourist assets. For example, the district government of Gianyar provides extra cash to the traditional village units (*desa adat*) to support custom-related (*adat*) activities.

The provincial and district governments also provide funding for education and other activities to maintain Balinese culture and language. The Congress of the Balinese language held every five years also gets government support. Teachers of primary schools are given extra monthly incentives, in addition to their salaries. Teaching materials for Balinese are developed and funded by the local Bali government.

NGOs and wealthy Balinese entrepreneurs also provide important support. There have been local newspapers, radios and televisions – Bali TV in addition to the local government TV- with programs in Balinese, addressing local issues. Balinese gets support from Internet technology, e.g. computerization of Balinese scripts in unicode (<http://www.babadbali.com/aksarabali.htm>).

There has also been a new movement called *Ajeg Bali* ‘Bali First, Prosperous Bali’. This movement aims at maintaining and reviving Bali in its original culture by empowering the Balinese people. It includes a range of community programs including business training and it provides small credit loans for the Balinese so that they can start up new businesses. The idea is that if the Balinese are economically better off, they will continue practicing and will help safeguard the Balinese culture.

In fact, for centuries, the Balinese have been able to maintain their unique culture, which is based on Hinduism. Bali Hinduism is a type of Hinduism that originated in India, reaching Bali in the 8<sup>th</sup> century A.D. It has incorporated many local cultures, including the Javanese culture. The Balinese follow long traditions of tight Hindu-based social structures. Hindu-based rituals and other practices are part of the daily life.

The traditional Hindu-based social structure has proven to be quite a resilient shield against the influence or pressure of outside cultures. Java (west of Bali) and Lombok (east of Bali) are now both Muslim islands. While the people in Java, Lombok, and other parts of Indonesia have either embraced Islam or Christianity, the majority of Balinese are now still embracing local Hinduism. Christian missionaries attempted to spread Christianity in Bali in the 20<sup>th</sup> century, but were unsuccessful.

There is world wide evidence that the spread of Christianity and Islam has resulted in the endangerment and ultimate extinction of local beliefs, related ritual practices, and finally local cultures, e.g. the case reported in the Trobriand Islands (Senft 1997; to appear). Certain Javanese dances and ritual practices using old Javanese (*Kawi*)—a language already extinct in Java - are still alive and are performed in Bali because they are related to the Hindu-based rituals still practiced in Bali but abandoned in Java. Bali is indeed a living museum of Hindu-Javanese culture (Soedarsono 1974:136).

The strength and resilience against the spread of Islam and Christianity can be attributed to the functional existence of the traditional customary village units (*desa adat*) in Bali. These customary villages are still strong, even in highly developed urban areas like Denpasar and Kuta. The uniqueness of *desa adat* is that it is not simply a territorial unit, but more of socio-religious unit (Warren 1993; Surpha 1995; Imawan 2003). For example, a single *desa adat* is united by the existence of three community temples, known as *kahyangan tiga*, namely Pura Desa, one Pura Puseh, and one Pura Dalem. The socio-religious concept of *kahyangan tiga* introduced in the 11<sup>th</sup> century by Mpu Kuturan (a priest from East Java) succeeded in unifying all sects of Hinduism in Bali. Mpu Kuturan also rearranged the structural concept of various shrines at all territorial levels, including the family, village and island levels.

Customary villages are led by a committee of community elders, headed by a *kelihan desa* or *bendesa adat*. Each *desa adat* has its own *awig-awig* (customary laws), specifying among other things proper conduct and obligations of community members, conflict resolutions and sanctions. The *awig-awig* laws are in Balinese; the ritual and community meetings are also in Balinese. While the *awig-awig* laws do not specifically stipulate that the language used must be Balinese, Balinese is the language that has been used, and will remain functionally so in these traditional domains for years to come. This will remain the case as long as the Balinese continue living with their traditional social values and structures.

It is indeed a challenge how to live in the modern world without abandoning ones traditional value, culture and language. The local government of Bali has taken initiatives to protect the Balinese customary (*adat*) system. For example, there have been local laws (*peraturan daerah*, or *perda*) issued by the provincial government to regulate the co-existence of *desa dinas* (the Indonesia administrative village unit) and *desa adat* (the tradi-

tional village unit). *Perda* 06/1986, for example, specifies the consultation and coordination between both *dinas* and *adat* leaders at the village level ( 2). Note that this legislation, which still applies now, was in fact already in place during Soeharto's era.

The local community at large is also aware of the challenge to have the right balance between maintaining traditions and living prosperously in modern Indonesia. Evidence of this comes from the active participation of the community in the current movement of *Ajeng Bali* and local NGOs, as discussed earlier. In addition, innovative responses of the traditional *adat* institutions are also observed. The traditional leadership at the level of the village unit appears to adapt to modern challenges by extending their responsibilities beyond

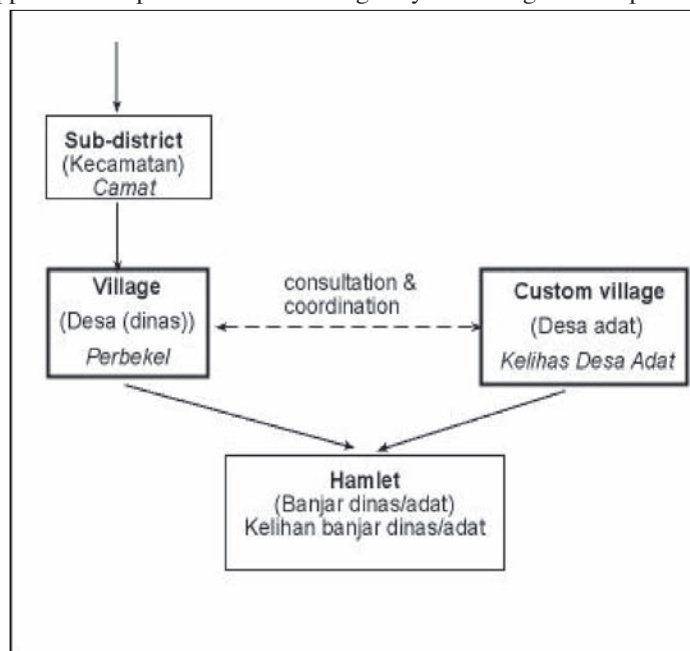


FIGURE 2. *Dinas* and *adat* village according to the Bali local law (*Perda*) 06 1986

*adat* matters, but still for the benefit of the *adat* community. While the village head (*kelihan adat*) and the members of the committee (*prajuru*) still exercise their traditional roles, they also give financial support to their local community. For example, many *adat* villages now run LPD (*Lembaga Perkreditan Desa*) or Village Credit Units providing loans and supporting businesses of the *adat* community members. Village cooperatives are also increasingly common now.

Certain villages, especially those in Kuta and Ubud, have done well, as they help manage tourism within their territories. The extension of this *adat* domain has resulted in the creation of jobs for the local *adat* members, and importantly good revenues for the *adat* village. The availability of self- and locally-generated revenues, autonomously managed by the village, allows the village to support local *adat* programs such as traditional rituals in temples, and public activities, and even to finance big village construction projects such as building community halls and village roads.

Excellent human resources also support Balinese. As mentioned earlier, the Balinese are in control of the local government in Bali from the provincial to village levels. There are tertiary institutions providing advanced studies on the Balinese language, art and culture: the Department of Local languages (Balinese and Old Javanese) at Udayana University, ISI Bali (Bali Institute of Arts), and Universitas Hindu Bali (Hindu University of Bali). Graduates of these institutions have worked for NGOs, local governments, and even have become independent business persons and artists that support and enrich Balinese arts and culture. As a result, creative arts and dances have blossomed in Bali in the last three decades or so.

There has also been a long tradition of literacy in local and modern scripts in Balinese. While traditional *lontar* (palm-leaf) writing and reading are still practiced, commercial books in Balinese in local and modern scripts are exploding in number, and available at local bookshops in Denpasar. In fact, a long time before the independence of Indonesia, the Dutch already set up Gedong Kertya in Singaraja (a town in northern Bali) which collected and documented Balinese *lontars*. Gedong Kertya has become the island's largest repository of *lontars*. In addition, the Faculty of Arts in Denpasar also has a special *lontar* library housing an incredible wealth of traditional knowledge.

To conclude, the provincial and district governments of Bali and the local community appear to have shown the capacity and experience necessary to manage their own needs, especially in relation to their local culture/language. The initiative of the local government to issue the local law (*perda*) regulating the *adat* village in 1986 is a genius move. Note that this came in the height of Soeharto's power and his campaign of Indonesianisation, dismantling or abolishing the traditional *adat* structures across Indonesia. It appears that Bali is now one place (if not the only one) where the local community (down to the village level) is ready for, has benefited from, and will continue to benefit from the new local autonomy granted by the central government in Jakarta.

**5.2.2 RONGGA IN FLORES.** My three-year ethnographic work in Rongga (2004-6) reveals quite a different situation. In contrast to Balinese, the local Rongga community suffers capacity-related problems. Discussion of these problems with reference to Balinese examples provides insights into the issues faced by disadvantaged minorities in modern Indonesia.

**a. Regional/Local Government Support.** Unlike Balinese, Rongga does not receive any support from regional and local governments. This is mainly due to regional/local politics in NTT (Nusa Tenggara Timur or East Nusa Tenggara Province) and in the district (kabupaten) of Manggarai where ethnic loyalty is strong. The total number of the Rongga people, which is around 5,000 compared with that of Manggarai (500,860),<sup>6</sup> is too small to be meaningful in Manggarai's politics. Given that numbers are important in modern democracy (as this determines who wins elections and controls the government) and given ethnic loyalty in local politics in Manggarai, the Rongga people are disadvantaged because there has not been, and will perhaps never be, a local government controlled by the Rongga people. All of the seven bupatis (regents) of Manggarai since the introduction of

<sup>6</sup> This is based on the figure in 2005 from Statistics of Nusa Tenggara Timur, <http://ntt.bps.go.id/pop/po01.htm>



the new administrative system in early 1970s have been Manggaraians, mainly from the Todo clan. The Todo people have traditionally dominated Manggarai since the seventeenth century when Manggarai was under the rule of the South Sulawesi kingdom of Goa (Erb 1999; Toda 1999). Moreover, there have been no Rongga people elected as members of the DPRD (Dewan Perwakilan Daerah, or Local Legislative Assembly). Even at the most local level, Kelurahan Tanarata (which is a Rongga village) is headed by a lurah who is not a Rongga. There is in principle nothing wrong with a non-Rongga person acting as the village head of a Rongga village if he can provide support needed for the local community. The fact is that, based on my observation in the field and discussions with the locals, the non-Rongga lurah does not live in the village, has no good knowledge of the local customary practices, and often cannot mediate or solve local problems.

In short, given the history of the local politics in Manggarai, any government controlled by the Manggaraian would not be expected to have any program with the best interest of the minority Rongga people.<sup>7</sup>

**b. Social Structure: the Role of Adat at the Grass Roots Level.** Balinese has its traditional *adat* structure intact, functional and dynamically responsive to new challenges. Rongga's traditional *adat* system, however, was demolished when Manggarai was turned into a district within the Indonesian administrative units in 1970. Subsequent changes at the village level including pemekaran (splitting into new villages) have further worsened the situation.

Historically, the district of Manggarai used to be divided into traditional unities incorporating around 39 principalities called *kedaluan*, headed by a *Dalu*. The old structure and its rough equivalence are presented in 3. The area of the Rongga community belongs to the traditional *kedaluan* of *Rongga-koe*. When the Dutch colonial government decided to take direct control over Manggarai in the early 20<sup>th</sup> century, the Dutch maintained the traditional *kedaluan* system and kept the political subordination of the Rongga people under the King of Todo. The traditional *kedaluan* system continued up to late 1960s, long after the independence of Indonesia in 1945. However, it was completely abolished when the New Order government of Indonesia introduced re-organisation to Manggarai in 1970, when Manggarai became a district. (Manggarai split into two districts in 2004.) There was also a change in the structure of village with the introduction of new Indonesian village administration (*desa gaya baru*).

As discussed in section 5.2.1, the introduction of a new Indonesian village administration in Bali does not result in the total replacement of the traditional *adat* village unit. Rather, it results in a dualism of *desa dinas* (Indonesian village administration unit) and *desa adat* (the traditional village unit). *Desa dinas* manages the official administration of the local government at the village level as an extension of the government structure,

<sup>7</sup> However, this might change given the current development of the formation of new districts in Manggarai. Manggarai has been split into smaller districts. It was split into Kabupaten Manggarai Barat (West Manggarai District) and Kabupaten Manggarai in 2003. A further split of Manggarai into Kabupaten Manggarai Tengah (Central Manggarai District) and Kabupaten Manggarai Timur (East Manggarai District) was officially approved by the central government in early 2007. Rongga villages are now part of Kabupaten Manggarai Timur. It remains to be seen whether this new development could indeed benefit minority communities like Rongga.

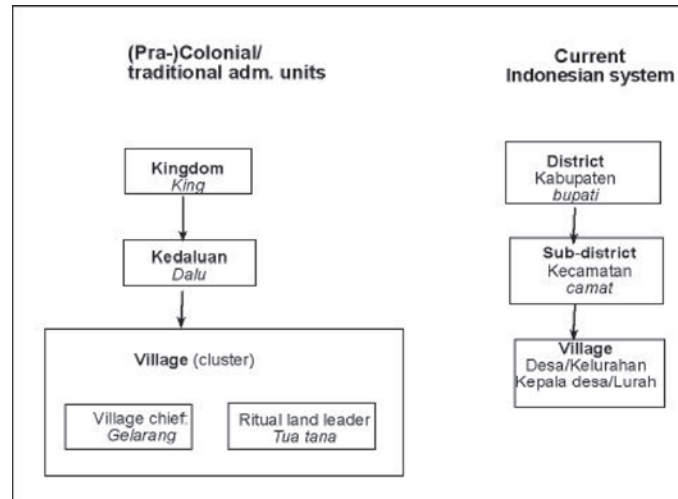


FIGURE 3. (Pre)colonial and current administrative units in Manggarai

whereas *desa adat* manages religious and other Hindu-based community affairs. Hence, the introduction of the new village administration does not have a negative impact on the local culture and traditional custom laws.

However, in case of Rongga, the introduction of new village administration turns out to have destroyed the traditional system that appeared to have worked well in the past in maintaining the integrity of the Rongga culture and language. In the past, there was a traditional system integrating or connecting groups of the Rongga scattered in the Rongga territories. This still worked under the *kedaluan* system, because Rongga was in one *kedaluan*, namely *kedaluan Rongga Koe*.

In modern Indonesia, however, the introduction of the new Indonesian administrative system has resulted in the collapse of this traditional local system. *Kedaluan* was totally removed. A number of *kedaluan* (which often used to be along the sub-ethnic grouping) were collapsed under a single new sub-district (*kecamatan*). An urban village administration system, particularly *kelurahan* (Javanese-based village systems), was introduced in the territory of Rongga. There has been a further split (*pemekaran*) or formation of new village administration units in the territory of Rongga. The traditional *adat* system and internal communication among the clans or groups of the Rongga people to address their own interests as a group has ceased to exist. Gone with the social structure is the traditional leadership, further discussed in (d) below. For example, when there was a dispute in 2003-4 between the local Rongga community and the local government of Manggarai regarding the status of land now owned by many Rongga families, the village officials (i.e. current *Lurah* and village heads) could not help. Mr. Anton Gelang, the former deputy of *Dalu* in the *kedaluan* Rongga Koe, told me that the people asked for his help. With his knowledge about traditional land rights and experience in resolving land disputes and with his former

role in the traditional leadership of *kedaluan*, he managed to resolve the problem with the government. He went to see the *Bupati* (Regent) in the capital town of Ruteng. Backed up by his personal detailed knowledge about the history and the related documents of land distribution in Rongga territories, he managed to convince the *Bupati* not to take over the disputed land.

**c. Human Resources.** The Rongga ethnic group is small (5000 people), far smaller than the Balinese (3 million people). We cannot therefore make a valid comparison. Nevertheless, it provides us with a window to understand how ‘being small’ in modern Indonesia is indeed a real problem.<sup>8</sup> Thus, the Rongga people do not control the local (sub)-district government, not even in their own village of Tanarata. Young Rongga people who have received higher education, generally at undergraduate level, are few. Most of them who do receive higher education do not go back home to their Rongga territory. When they do, they are often unemployed, or have to do small farming.

Few educated young Rongga people are committed to the maintenance, let alone revival, of Rongga. Only three community elders appear to have a strong desire to do something for their culture and language: two are retired teachers and one is a former deputy *Dalu* (all in their sixties). The two retired teachers have done some documentation of their own, collecting traditional stories and ritual texts written down in handwriting in note books. One of them also taught traditional Rongga songs and dances to primary school students. They admit that they cannot do much because of health and economic problems. They promised that they would help and were emotionally moved when they learned about the goals of my language documentation project. I have already recruited one Rongga person (my research assistant), and trained him to perform language documentation. These people seem to be motivated to do something for their language and culture. There are certainly not enough skilled people for a successful language maintenance program. In addition, it is a challenge to keep their motivation alive (for example my project is now over, and there is no more financial support for these activities).

**d. Leadership.** Traditional leadership of *kedaluan* (3) for the whole Rongga group is totally extinct now. While descendants of *daluh* (chief of *kedaluan*), *glarang* (village chief), and *tua tana* (ritual land leader) are still there, they are not functional. These traditional leaders can no longer exercise power to mobilize people or resolve local conflicts (particularly land disputes) as in the old days. Customary (*adat*) laws are generally no longer strictly practiced or socially imposed. Note that in Bali *adat* laws are generally still respected and enforced by the social structure.

Traditional leadership tied to the traditional social structure for the whole ethnic group is important for the survival of the culture and language of the ethnic group against external pressures or influences. This has been evident from the functionality of the *adat* village

---

<sup>8</sup> However, it should be noted that ‘small numbers’ in other socio-geographical and political contexts may not be a problem and may not constitute a variable contribution to the language marginalisation or endangerment. It has been reported that a small number of 500 speakers may turn out to be a healthy number, e.g. in the context of Pacific languages (Grimes 1995). Other factors such as the speaker’s attitude and external or macro-level variables (i.e. national, regional and local settings) are also crucial (Grenoble and Whaley 1998).

in Bali. As discussed earlier, the traditional system provides a means or a mechanism for the local community not only to take care of their own traditional internal matters such as community rituals and land disputes, but also to consolidate collective efforts to safeguard their interests against other ethnic groups at the local or regional levels. Unfortunately, this is lacking with the Rongganese as a result of the dismantling of the traditional socio-structural system of *kedaluan*. The Rongga people now appear to have become too occupied in internal conflicts and rivalries among clans, and are not fighting for their common interests at the district or sub-district levels in modern Indonesia.

In addition, we have observed how the *adat* structure in Bali can be innovatively extended to achieve or fulfill common economic benefits. The abolishment of *adat* structure at the village level in Rongga has led to the unavailability of means for the local community elders/leaders in Rongga to function as well as the Balinese village leaders. While there are Rongga clan chiefs around, they are not in any way part of *adat* organization at the village level, nor are they elected by the community members, and therefore they have no authority in leading the community at the village level. Unfortunately, the current village chiefs, who are not clan elders, are rather passive (as is the case with civil servants in Indonesia generally). They are in a position to lead in the absence of functional *adat* structure in Rongga, and could play a key role. However, this has not happened (yet).

**e. Funding Support, Local Economy and Poverty.** Successful programs need good financial support. Therefore, a strong regional and local economy is an important factor for cultural and language maintenance or revival. It is not cheap to run and support long term language/cultural maintenance and revival programs. The costs include expenses for providing the necessary training of the local language workers and for building a local language and cultural centre. This centre also must have facilities that enable the local language workers to work with the local community. Maintaining the facilities and providing support for the programs run by the language workers (e.g. production of materials for literacy programs) may run to millions in Indonesian rupiah annually, which is expensive in local terms. The provincial and district governments in Bali, particularly the districts of Badung and Gianyar, could provide financial support to a range of local cultural and language programs including *desa adat* because of good local revenues based on local taxes and other locally-generated government revenues or PAD (*Pendapatan Asli Daerah*), under the local autonomy.

The regional/local economy in Flores and NTT (East Nusa Tenggara) province is a different story. NTT is one of the poorest regions in Indonesia. The local district government of Manggarai does not have much of a budget for cultural community programs for ethnic minorities like Rongga. While there is some funding from the central government, it is to support basic needs, e.g. distribution of *raskin* (*beras miskin* or rice for the poor) and cash for the poor. Even in this case, from my experience in Rongga, much of the money often does not reach the ordinary people who so desperately need it, as it encounters corruption along the way, even by the village head. It is well known that corruption is indeed a serious problem across Indonesia.<sup>9</sup>

<sup>9</sup> Corruption of government assistance for the poor such as *raskin* (rice for the poor), JPS (*Jaring Pengaman Sosial* or social security net), BLT (*Bantuan Tunai Langsung*, or direct cash assistance) has been well reported in the media in Indonesia, e.g. the corruption case in Bogor (West Java) in

Recall that there is active community/private participation and support in cultural/language revival programs in Bali, e.g. the *ajeg* Bali movement, the launch of private Bali TV, palm-reading/singing groups for shows or competition on local TV stations. This active participation and support is made possible because the indigenous economies of Bali are strong. Ordinary people are by and large doing well economically in Bali. Only 6.8 % of Balinese are below the poverty line (base on BPS 2006).

In NTT, in contrast, 77% of the people are living below the poverty line. My field-work experience in Rongga villages suggests that the percentage is even higher than this official statistic. I noticed that most of the ordinary people struggled to have daily meals. Naturally, they did not enthusiastically participate in cultural programs since getting food was their priority. It is often the case that people expect to be paid for their participation in community programs. This can be perhaps partly be blamed on the common practice of the new order government, that if there is a government project, money from the government is involved, and they expect to have a share of it. In Bali, in contrast, people would be generally willing to donate and spend their own money for certain festivals/programs.

Thus, there is a clear link between the motivation for cultural/language maintenance and the state of local economies. Edward (2004) highlights the importance of indigenous economies in language maintenance. He points out that local business initiatives offer local communities opportunities, independence, and power to conduct their affairs in accordance with their traditional values. In addition, the local workplaces created by good local economies are a natural setting for using (minority) local languages. Comparing the local economies where Rongga and Balinese are used shows how minorities are disadvantaged, helpless, and not enthusiastic in cultural/language programs. In addition, it also points to the difficulty in empowering them, because getting rid of poverty, or alleviating it, is a complex issue that has no easy solution.

**f. Literacy, Literary Tradition and, Language Teaching.** Writing provides another means by which traditional knowledge and other cultural information can be stored and transmitted across generations. The written medium also enables creative literary work that in turn enriches the culture, and contributes to the viability of the culture and language. Teaching local children the ability to read and write in their local language, in addition to speaking it, is as important as the written tradition because it will ultimately enhance the capacity of the children to help maintain their language and culture.

Balinese also has a long tradition of literacy in local (palm leaf or *lontar*) and modern scripts. Short stories, novels and poems have been published in Balinese since the early 20<sup>th</sup> century. Rongga in contrast has no similar local script. While there has been some work on poems, they are not yet published, but only produced in a restricted context in church. The closest thing to traditional poems in Rongga is *vera*. However, *vera* is an oral verbal art performed simultaneously with dancing. *Vera* is never written or produced in written form.

the daily *Kompas* (<http://www.kompas.com/kompas-cetak/0505/04/metro/1726898.htm>). A detailed discussion of key aspects of wide spread corruption in Indonesia is given in Henderson and Kuncoro (2004).

The standard orthography of the local language is important in literacy programs. Balinese has long had an established orthography. Rongga orthography is, however, still in the process of being standardized. I have initiated this as part of the Rongga documentation project, however approval from the wider Rongga community is needed.

Including local language in the school curriculum is also an important part of a literacy program. In Bali, Balinese is the local language taught as part of the *mulok* (*muatan lokal*, or local curriculum content) to all students of primary schools in Bali, even to the non-Balinese who happen to attend public schools in Bali. In contrast, Rongga is not taught even at the primary schools in the Rongga territories. The chosen local language selected as *mulok* is Manggarai, the language of the dominant ethnic group of the district. Again this highlights the problem of how minority languages are disadvantaged even at the district level of local autonomy.

Developing teaching materials also poses a problem for minority languages due to cost and time constraints. Languages selected by the local government as *mulok*, such as Balinese and Manggarai, receive a special budget for this from the district government. Teaching materials are therefore generally well developed and made available to local teachers for free. In contrast, teaching materials for a minority language like Rongga are not available, and have to be personally developed and produced by the local teacher. It is often the case that he/she has not much time to do this.

Furthermore, there may be problems in teaching the materials at the local school. In collaboration with a local school teacher, I developed teaching materials in Rongga. However, there was a problem of finding the time slot in the already over-crowded schedule. In addition, there was another problem: the ‘no benefit’ problem with respect to student performance in the *mulok* test. Since the *mulok* test is in Manggarai, it makes sense to give Manggarai priority, not Rongga. While there is no evidence to support this, the primary school teacher I talked to was concerned that giving priority to Rongga would mean less time for teaching Manggarai. And, according to them, this might result in students’ poor performance on the *mulok* test.

**6. FINAL NOTES.** In this final section, I provide a brief summary, followed by discussions of capacity building, long-term support, and future prospects.

**6.1 SUMMARY.** The summary in Table 1 shows how Rongga (column 3) is disadvantaged in all aspects that are enjoyed by Balinese (column 2). This appears to be a typical picture of a small language of a marginalized minority ethnic group. Rongga lacks institutional, organizational and financial support at the regional/provincial level as well as at the district and sub-district levels. Regional and local politics do not help either. The function of the traditional *adat* structure is diminishing, and totally disconnected from the village (social) structure. This has long crippled the *adat* laws and has not been healthy for the maintenance of the culture and language. In addition, widespread poverty is a big problem. This has resulted in poor human resources and also the inability of the local people at the grass roots level to actively participate in programs organized by outsiders and/or to be independently engaged in programs they design themselves.

**6.2 CAPACITY BUILDING: WHAT, WHO, AND WHERE?** Given the complexity of cultural, socio-political and economic variables involved in the local levels in the maintenance and revival of local languages as discussed earlier, capacity building must address issues related to these variables at different levels. To be useful and realistic, it is important to adopt the view that capacity building is more than simply developing human resources. Capacity building must also include strengthening organizations and reforming institutions (cf. Grindle 1997). I suggest that these must be at national, regional and local levels. This is depicted in 4.

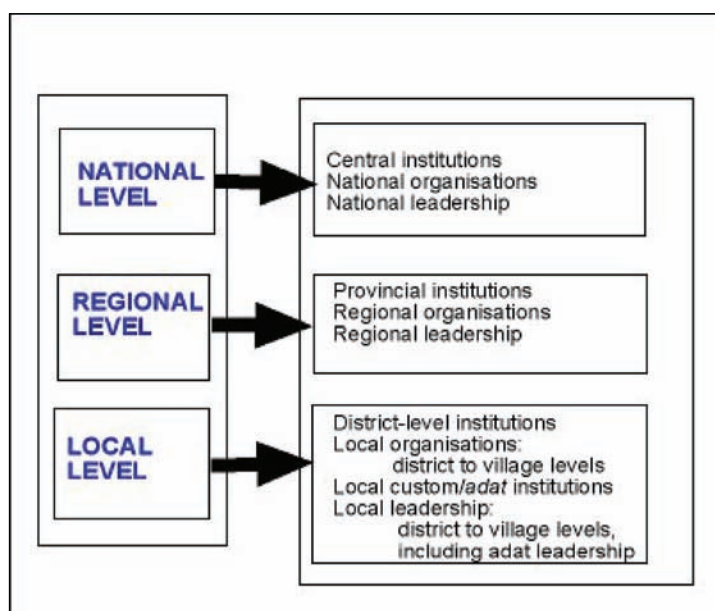


FIGURE 4. Targeted institutions, organisations, and groups in capacity building programs<sup>10</sup>

It is clear that the task of capacity building in this view is a huge one. However, there are certain things that a field linguist can do to help (discussed below). Of course, there are complex matters that are beyond the capability of an ordinary field linguist, e.g. alleviating poverty (see the cells with a question mark in Row 4 of Table 1 for a complete list).

A linguist can certainly help to develop human resources for language maintenance. Ideally, this must be done at all levels. However, the emphasis must be on developing human resources at the regional, district and village levels as these are the levels that actually matter and could make a difference. The options include short training courses, training of trainers, enrolment in academic programs, and consultancy services. The choice of programs depends on available funding. However, it is perhaps realistic that linguists and other faculty members of the regional/local university, e.g. those in Ruteng or Kupang in case of NTT are targeted in this capacity building program. These faculty members are

<sup>10</sup> This model is inspired by Mera (2000).

respected and listened to by the locals. They are most likely to be consulted by the local government and are in the position to apply for grants from the regional and/or central government. Such grants then can fund research and community programs that include the local people at the grass-roots level.

My experience from Flores suggests that local teachers and *tokoh adat* (customary elders) may play a significant role and therefore must be targeted in capacity building programs. A capacity building program targeting mixed groups of Indonesians (university lectures, teachers, and language workers) from across the archipelago was initiated in 2006 in the form of a language documentation workshop, in which I took part (see Florey and Himmelmann 2008). As part of Rongga documentation project, I also organized two workshops (2004, 2006) at Udayana University in Bali targeting university lecturers from across Indonesia.

For the basis of the discussion in this paper, I suggest that capacity building must also include reviving the role of traditional customary (*adat*) laws at the village level. New autonomy laws 22/1999 and 32/2004 provide a legal basis for reviving and strengthening the customary system. From the foregoing comparative discussion of Balinese and Rongga, it should be clear that strengthening institutions/organizations (including *adat* ones) is as important as developing human resources. We have seen how the traditional social structure at the village level of *desa adat* in Bali is still intact, functional, and can adapt to new challenges, e.g. revising *awig-awig* (*adat* laws) and extending its responsibilities to include managing modern businesses to provide financial support for local community activities.

Reviving indigenous *adat* structures would not be an easy task for Rongga. *Adat* in Rongga is increasingly restricted to ceremonial or ritual affairs (*upacara adat*) in relation to *rumah adat* (the clan house) and the garden. There has been an attempt recently to form a committee of elders in Rongga headed by the former Deputy Dalu, Mr. Anton Gelang. However, this does not appear to go far enough to revive the social structure that used to be functional for the Rongga ethnic group. Modeling on the dualism or co-existence of Balinese *dinas* and *adat* at the village level, and forming another upper layer of structure of *adat* for the whole ethnic group could be an alternative to explore.

The significance of reviving (or maintaining) custom and customary leaders gets empirical support from other parts of Indonesia. In Kei (south-east Maluku), for example, traditional leadership within traditional local autonomy has proven to be able to stop Christian-Muslim fighting spreading from Ambon to the island (in 1999) and brought the community together again (Laksono 2002; Laksono and Topatimasang 2004). While the earlier law on village government (5/1979) had totally destroyed traditional village autonomy, one of the customary Kings of Waltar, Raja JP Rahai, succeeded in retaining customs in his territory which consists of several villages with different religions.

When the conflict broke out, using his traditional authority and prestige, he managed to stop the conflicts, not only within its territory but also in other villages as he succeeded in launching a movement of customary reconciliations with other traditional leaders. He inspired Kei Islanders with the idea that they belonged to one community long before the segregation created by the arrival of world religions (Islam, Catholic and Protestant), which often coincide with current Indonesian administrative units of villages, and that peace depended on community initiatives. This traditional concept of 'one-ness' played a significant role in ending the conflict in Southeast Maluku. Arguably, there always appears



to be a similar indigenous concept in other cultures, e.g. the concept of *krama* Bali (roughly, Balinese brotherhood) in Balinese. This traditional concept can be revived, not only to resolve local conflicts, but also to safe-guard local traditions, cultures and languages.

**6.3 LONG-TERM SUPPORT.** Capacity support is as important as capacity building. Institutions, organizations, and legislatures can be reformed, created, or strengthened at a particular point in time. However, long term support is needed to ensure that whatever has been reformed or created functions well and that the functionality is sustainable. Providing long-term support for local communities highlights the need for developing good human resources at the most local level, because they are agents of change who live with, or are geographically close to, the local community. Geographical proximity would enable them to have regular or constant interaction with the local community. In this way, support for the local needs can be delivered efficiently. This is critical since communication and traveling is generally a problem in eastern Indonesia.

The availability of local institutions, infrastructure and incentives for those who have been trained so that they can put their skills and knowledge into practice is important for long-term support. At the national level, international organizations such as the United Nations, the World Bank, and the Asian Development Bank have been working on the issue of capacity building (for purposes more broad than simply cultural/language maintenance).

However, young Indonesians who are highly trained overseas in advanced countries are unable to play a role in managing and advancing their own country because of lack of infrastructure, incentives and institutional/organisational support. The same is true at the local level in Flores generally. Young locals educated outside Flores (mainly in Java or Bali) are back home but they are unable to work in their field of training, or they are unable to participate in governance. There are at least three locals who have been trained for language documentation in my Rongga project. However, the unfortunate fact is that they have not been able to do language and cultural documentation without my continuous financial support.

Long-term support in terms of infrastructure and incentives is ultimately the responsibility of the local government. The officials of the relevant local government institution must therefore be targeted in any local capacity building program. Their awareness of the issues involved and their active involvement and support is crucial.

**6.4 PROSPECTS.** Both capacity building and support involves complex and inter-related variables as seen in Table 1. While there appears to be no easy shortcut to solutions, and the desired outcome can not be achieved in the foreseeable future, current political change towards decentralization in Indonesia brings about a good prospect for reviving the local cultures and languages. The challenge now is how to turn it into reality, especially for disadvantaged minority groups like Rongga. It should be noted, however, that the national and local politics in fast-changing Indonesia has led to uncertainties whether the central government and DPR would whole-heartedly implement local autonomy as in the spirit of reformation, and whether priority would be indeed be given to local language/cultural maintenance.

Generally speaking, an ethnic group that is big enough to control the entire district government has the opportunity to take full advantage of local autonomy in modern Indonesia because autonomy is given to the district government. Controlling the district government means controlling policies and resources, including those that benefit their local languages/cultures. There is now a growing trend for *pemekaran* (splitting administrative units) across Indonesia. In Flores territorial boundaries of newly formed district units often coincide with the territorial boundaries of ethnic groups. For example, the former district of Ngadha, which used to include the areas of Ngadha and Naga-Keo, now has been split into the new districts of Ngadha and Naga-Keo. In theory, this trend would benefit local cultures and languages of the controlling ethnic group in the district.

Unfortunately, it appears that small minority groups such as Rongga would still be disadvantaged because they are scattered in a couple of villages, too small in number to control the district government. The district government is, and will always be, controlled by the majority group, e.g. in the case of Rongga in Manggarai, by the Manggaraian. One way of getting around this problem is to revive and strengthen traditional customary (*adat*) autonomy at the village level.

As suggested earlier, the Bali model of dualism and co-existence of village social structures could be used as a model as this co-existence has proven to benefit local cultures/languages. However, this requires specific local legislation and therefore the good will and cooperation of the dominating ethnic group to pass such legislation. Reviving the local customary system would ensure that the minority groups can take care of their interests, and that there is no return to Soeharto's style of controlling and politicizing villages for the benefit of certain groups within the district government.<sup>11</sup>

Fortunately, Indonesia appears to have reached a point of no return in terms of democracy. Ethnic groups (both large and small) have formed powerful alliances called AMAN (*Aliansi Masyarakat Adat Nusantara* or the National Alliance of Customary Societies) whose goals, among others, are to revive and empower the *adat* customary systems including the minority ones. Importantly, they have encouraged *tokoh adat* (prominent *adat* elders) to get actively involved in local, regional, and national politics, and to run in the election for the local legislative assembly (DPRD) and central legislative assembly (DPR). Their active participation within a modern democratic system would hopefully lead to a greater number of members of DPR(D) elected from educated ethnic elders. Then, hopefully bio-cultural diversity of indigenous people of Indonesia with the associated *adat*/customary rights would be well taken care of in New Indonesia. There is a downside of this, however. *Adat* could become a political commodity exploited for the personal benefit of the elders rather than the community.

---

<sup>11</sup> The return of 'centralisation' of the Soeharto's style under the guise of district autonomy is possible, given the fact that the district government is powerful under the new autonomy.

TABLE 1 - Balinese and Rongga compared: a summary of a range of inter-related variables involved in capacity building and support

(1)	(2) Balinese	(3) Rongga	(4) Notes; what to do?
Institutional/ organizational support:			
<ul style="list-style-type: none"> <li>Local Government &amp; Institutional support</li> </ul>	Strong at provincial and district levels	Almost no support	Complex issues involved
<ul style="list-style-type: none"> <li>Traditional social institution/ structure: customary (<i>adat</i>) system</li> </ul>	Strong at the village level; co-existence of official ( <i>dinas</i> ) and traditional customary systems	The <i>adat</i> system almost ceased to be functional at the village level	Revive the customary system; local government legislations ( <i>perda</i> ) are needed
<ul style="list-style-type: none"> <li>Community support</li> </ul>	Strong Bali wide	Weak	Complex issues involved: increase awareness and empowerment of the locals
<ul style="list-style-type: none"> <li>Non-Government Organisations (NGOs)</li> </ul>	There are NGOs dealing with language and cultural matters	Available NGOs generally deal with poverty alleviation.	Encourage local NGOs to extend activities include cultural and language matters
Financial support:			
<ul style="list-style-type: none"> <li>Regional/local economy</li> </ul>	Good	Poor	Complex issues involved, almost beyond anybody's control
<ul style="list-style-type: none"> <li>Government funding for cultural/ language programs</li> </ul>	Good contribution by provincial and district governments	very little or none	The situation could be better in future under local autonomy

<ul style="list-style-type: none"> <li>• Non-government funding for cultural/language programs</li> </ul>	NGOs, community, individuals	none	Help to set up local NGOs; help them apply for grants for local projects
<ul style="list-style-type: none"> <li>• Average income and poverty</li> </ul>	Bali: 6,8% below poverty line	East Nusa Tenggara: 52% very poor 25% poor (in 2002); 27,86 % below poverty line	Complex issues: widespread poverty is a real problem in NTT
Regional and local politics	Provincial and district levels controlled by the Balinese	Disadvantaged: not in control of the local government, even at the sub-district (or village) level	Encourage local elders/leaders to run as candidates for local DPR? Seems not realistic at least now
Literacy and education	Long literacy tradition in local and modern scripts; Balinese taught as <i>mulok</i> (local curriculum contents) at schools; Department of local languages at Udayana University; commercial books in/about Balinese available at local shops in Bali	No literacy tradition in Rongga; Rongga is not taught as <i>mulok</i> even at local primary schools.	Help produce books, stories, teaching materials etc.; standardize orthography; lobby and work with the local education department to develop and teach <i>mulok</i> for Rongga

Literary tradition	Long literary tradition	Almost no literary tradition	Help publish books, stories, literary work by the local authors.
Human resources	Good	Poor	Short training; enrolment at local university; consultancy
Leadership:			
<ul style="list-style-type: none"> <li>Traditional leadership</li> </ul>	Functional at the village level; proven to be a crucial factor	Almost ceased to exist at the village level.	Help revive the local customary <i>adat</i> system; adopt and the co-existence model of official ( <i>dinas</i> ) and <i>adat</i> system as practiced in Bali
<ul style="list-style-type: none"> <li>Current context of modern democracy and autonomy in Indonesia</li> </ul>	Benefited: the Balinese are the majority at the Province and district levels in Bali	Disadvantaged: Rongga is a minority group even at the district level	Join the national alliance of <i>adat</i> societies (AMAN)

## REFERENCES

- ARKA, I WAYAN. 2005. Challenges and prospect of maintaining Rongga: a preliminary ethnographic report. In *Proceedings of the 2004 Conference of the Australian Linguistics Society*, edited by I. Mushin. <http://deneb.library.usyd.edu.au:8080/handle/2123/138>.
- ARKA, I WAYAN. 2006. Balinese. In *Encyclopedia of Language and Linguistics*, edited by K. Brown, 654-657. Oxford: Elsevier.
- BJORK, CHRISTOPHER. 2004. Decentralisation in education, institutional culture and teacher autonomy in Indonesia. *International Review of Education* 50:245-262.
- DE GRAFF, TJEERD, AND HIDETOSHI SHIRAIISHI. 2004. Capacity building for some endangered languages of Russia: Voices from Tundra and Taiga. In *Language documentation and description, Vol. 2*, edited by P. A. Austin, 59-70. London: Hans Rausing Endangered Languages Project.
- DIMMENDAAL, GERRIT J. 2004. Capacity building in an African context. In *Language documentation and description, vol 2*, edited by P. A. Austin, 71-89. London: Hans Rausing Endangered Languages Project.
- EDWARD, VIV. 2004. *Multilingualism in the English-speaking world*. Oxford: Blackwell.
- ERB, M. 1999. *The Manggaraians: A guide to traditional lifestyles*. Singapore: Times Editions.
- FLOREY, MARGARET, and NIKOLAUS P. HIMMELMANN. 2008. New directions in field linguistics: Training strategies for language documentation in Indonesia. In *The endangered languages of Austronesia*, edited by Margaret. Florey. Oxford: Oxford University Press.
- FOLEY, WILLIAM A. 2004. Language endangerment, language documentation and capacity building: challenges from New Guinea. In *Language documentation and description, vol 2*, edited by P. A. Austin, 28-38. London: Hans Rausing Endangered Languages Project.
- GRENOBLE, L.A., and L.J. WHALEY. 1998. Toward a typology of language endangerment. In *Endangered languages: current issues and future prospects*, edited by L. A. Grenoble and L. J. Whaley, 22-54. Cambridge: Cambridge University Press.
- GRIMES, J.E. 1995. Language endangerment in the Pacific. *Oceanic Linguistics* 34 (1):1-12.
- HENDERSON, J. VERNON, and ARI KUNCORO. 2004. Corruption in Indonesia. Paper presented at NBER Working Paper No. 10674, <http://www.nber.org/papers/w10674>.
- IMAWAN, RISWANDHA. 2003. Catatan kritis pelaksanaan otonomi tingkat desa di Bali. In *Kompleksitas persoalan otonomi daerah di Indonesia*, edited by A. G. Karim, 327-244. Yogyakarta: Pustaka Pelajar.
- LAKSONO, P.M. 2002. We are all one. *Inside Indonesia*, April-June 2002.
- LAKSONO, P.M, and ROEM TOPATIMASANG, eds. 2004. *Ken sa faak: Benih-benih perdamaian dari kepulauan Kei*. Tual-Yogyakarta: Nen Mas II - Insist Press.
- LASTRA, YOLANDA. 2004. The need for capacity building in Mexico: Misión de chichimecas, a case study. In *Language documentation and description, vol 2*, edited by P. A. Austin, 108-121. London: Hans Rausing Endangered Languages Project.
- MERA, KOICHI. 2000. Decentralization and capacity Building: Selecting Modes of Training for Indonesia. Paper presented at Presented at 47th North American Meetings of the Regional Science Association International, Chicago, November 9-12, 2000.

- SAID, M MAS'UD. 2005. *Arah baru otonomi daerah di Indonesia*. Malang: UMM Press.
- SENF, GUNTER. 1997. Magic, missionaries and religion: some observations from the Trobriand Islands. In *Cultural dynamics of religious change in Oceania*, edited by T. Otto and A. Borsboom, 45-58. Leiden: KITLV Press.
- SENF, GUNTER. to appear. Culture change - language change: Missionaries and moribund varieties of Kilivila. In *Endangered Austronesian, Papuan, and Australian Aboriginal languages: Essays on language documentation, archiving, and revitalisation*, edited by G. Senft. Canberra: Pacific Linguistics.
- SIL INTERNATIONAL, INDONESIA BRANCH. 2001. *Languages of Indonesia*. Jakarta: SIL International, Indonesia Branch.
- SNEDDON, JAMES. 2003. *The Indonesian language: Its history and role in modern society*: UNSW Press.
- SOEDARSONO. 1974. *Dances in Indonesia*. Jakarta: Gunung Agung.
- STEINHAEUER, HEIN. 1994. The Indonesian language situation: prospects and possibilities. *Bijdragen tot de Taal-, Land- en Volkenkunde* 150 (4):755-783.
- SURPHA, I WAYAN. 1995. *Eksistensi desa adat di Bali*. Denpasar: Upada Sastra.
- TODA, DAMI N. 1999. *Manggarai: mencari pencerahan historiografi*. Ende: Nusa Indah.
- TURNER, MARK, OWEN PODGER, MARIA SOMARDJONO, and KTIRTHAYASA. 2003. *Decentralisation in Indonesia*. Canberra: Asia Pacific Press.
- USMAN, SYAIKHU. 2002. Regional autonomy in Indonesia: field experiences and emerging challenges. Paper presented at The 7th PRSCO Summer Institute/The 4th IRSA International Conference: "Decentralization, Natural Resources, and Regional Development in the Pacific Rim", Bali, 20 - 21 June 2002.
- WARREN, CAROL. 1993. *Adat and dinas: Balinese communities in the Indonesian state*. Kuala Lumpur: Oxford University Press.

I Wayan Arka  
wayan.arka@anu.edu.au

# 6

## Documenting and Revitalizing Kavalan

Fuhui Hsieh and Shuanfan Huang

*Tatung University and National Taiwan University*

The purpose of this paper is to provide a two-dimensional approach to language documentation (Himmelmann 1998). In addition to building a database, we also conducted a sociolinguistic survey designed to document the state of health of a language in a particular spatio-temporal frame. Our goal is to share our fieldwork experience of documenting Kavalan, a seriously endangered language in southeastern Taiwan now spoken by fewer than just a few dozen speakers. We first discuss our field experiences in working with speakers of Kavalan in Sinshe village, the only significant Kavalan settlement left in Taiwan, and the state of the Kavalan language, based in part on Huang and Chang's (1995) earlier sociolinguistic survey, and in part on a recent more in-depth village-wide survey of language use in the community. Next, we introduce the NTU Corpus of Formosan Languages, part of which incorporates our corpus data in Kavalan. The NTU Corpus of Formosan Languages aims to establish a standard for the creation of linguistic corpus databases through the application of information technology to linguistic research. The creation of this linguistic database enables us both to preserve valuable linguistic data and to provide a systematic recording of these languages, for the benefit of future linguistic research.

**1. INTRODUCTION.** As the world's languages are dying at an unprecedented speed, language documentation has now been widely recognized as an important aspect of linguistic research.<sup>1</sup> As many as half of the estimated 6,000 languages spoken in the world are 'moribund' (Krauss 1992); in other words, these languages are spoken by adults only and are not being passed on to the next generation (cf. Crawford 1995). It has been estimated that ninety percent of the existing languages today are likely to die or become seriously endangered in the near future; and this phenomenon is particularly acute in Americas, Africa, Australia and Southeast Asia (Brenzinger 1992; Robins and Uhlenbeck 1991; Schmidt 1990). For most of the Austronesian languages spoken in Taiwan, the language crisis is imminent, and that makes documentation of Formosan languages all the more urgent. According to a most recent census report of the Council of the Indigenous Peoples in Taiwan, in August 2007, of the twelve indigenous tribes in Taiwan, there are five tribes whose populations number less than ten thousand each: Tsou (6,432), Saisiyat (5,514), Yami (3,125),

---

<sup>1</sup> We would like to thank the head of the Development Association of the Sinshe Community, Mr. Yinhua Pan (潘銀華), Mr. Jinlong Pan (潘金龍) and all the people we met in Sinshe for their help to make our field trip an enjoyable experience. We are also grateful to Professors Mingyi Wu (吳明義) and Shihhui Lin (林蒔慧) for their enthusiastic help to find Kavalan seniors in Sincheng Township for our language survey. Finally, we thank Ms. Fengyuan Yeh (葉鳳園) and Yuyuan Fang (方渝苑) in the Administration of Residents and Residence of Fengbin Township, Hualien for their assistance in providing us with valuable information we needed. The research reported here was supported by a research grant from National Science Council to the second author.



Kavalan (1,078) and Thao (619).<sup>2</sup> Their languages are long known to be at varying degrees of decline and thus potential extinction driven by the twin forces of rapid urbanization and Sinicization. Competent speakers of Kavalan, for example, are now estimated at less than just a few dozens.

The purpose of this paper is to provide a more in-depth approach to language documentation. In the words of Himmelmann (1998:165), “language documentation may be characterized as radically expanded text collection”, and its purpose is to represent the language for both linguists and the uninitiated, who do not have access to the language itself. Although text collection may help preserve (some aspects of) a language itself, it cannot reveal the true picture of actual language use by a community of speakers in a particular spatio-temporal setting. As a language is in decline, one may want to know what the various forces are which may be contributing to its endangered status, how the language is losing ground in the home and the work domains. A sociolinguistic study, appropriately designed, may certainly help provide, at least in part, answers to these questions.

In this paper, we report our field experience of documenting Kavalan, a seriously endangered Formosan language spoken in southeastern Taiwan. In section 2 we give a brief introduction to the history of the Kavalan people. In the second part of the section we report on our recent field trip to Sinshe Village and present some of the data regarding the state of health of the Kavalan language in this village. In Section 3, we introduce the NTU (National Taiwan University) Corpus of Formosan Languages, which is built with an attempt to establish a standard for the creation of linguistic corpus databases through the application of information technology to linguistic research. As pointed out by Lehmann (2001:87), “an important specific purpose of language documentation is to serve as a record of the past and as an element of ethnic identity for future members of the community that has lost its identity as a speech community but which still recalls that their ancestors had a language of their own.” The creation of this linguistic database thus enables both experts and common users to preserve valuable linguistic data and to provide a systematic recording of these languages.

**2. CURRENT STATE OF THE KAVALAN LANGUAGE.** In this section, we will first give a brief introduction to the history of the Kavalan people. Then, we will report on our recent sociolinguistic survey conducted in Sinshe Village and present some of the data regarding the state of health of the Kavalan language in this village.

**2.1 A BRIEF HISTORY OF KAVALAN.** For many centuries, the Kavalan people inhabited the present-day Ilan area in northeast Taiwan, which was known as *kap-a-lan* (蚵仔難 or 甲仔難), a transliteration of the word ‘Kavalan’. These people called themselves Kavalan, meaning “people living in the plains”, to distinguish themselves from the other aboriginal people in the mountain areas, e.g., the Atayals. Into this fertile land of tranquility toward the end of the eighteenth century came hordes of Han Chinese, which soon set in motion a series of arduous and probably sometimes heart-rending southward migration by the Kavalan people.

---

<sup>2</sup> According to the census of the Council of the Indigenous Peoples in Taiwan, the total population of the indigenous peoples in Taiwan in August 2007 is 481,119, which is less than 2% of the population of Taiwan. <http://www.apc.gov.tw/chinese/docDetail/>

The first migration took place during the period between 1830 and 1840. In 1796, the first group of the Han Chinese, led by Wu Sha (吳沙), moved in and opened up their first settlement in Toucheng (頭城).<sup>3</sup> As more and more Han Chinese followed into the area and took over, by force and by craft, the Kavalans' land, a number of sporadic migrations occurred, principally to Sanshing (三星) and Suao (蘇澳). Between 1830 and 1840, as they were losing their land and thus their socio-economic dependence, the Kavalan people, led by Kaliwan tribe living in Tongshan Township (冬山鄉), underwent a massive southward migration, and took up residence in Sincheng, Hualien (新城,花蓮), where their settlement was known as Kaliwan Village (加禮宛社).<sup>4</sup>

The Kavalan people did not stay there for long because of the Kaliwan Incident (加禮宛事) in 1878. There are two different versions of what triggered the Kaliwan Incident. One version holds that the Han Chinese, led by a businessman Wenli Chen (陳文禮), invaded and took over the land belonging to the Kaliwan Village. Another version holds that the Qing official Huihuang Chen (陳輝煌) swindled the Kaliwan people out of lots of money, and people in the Kaliwan Village, aided by Sakizaya, rose up to fight against the Qing soldiers. Many Kavalan people died in the battle. Fearing possible retaliation by the allied power of the aboriginal peoples, the Qing government forced the Kavalan people to move out of the area. This forced migration resulted in most of the remaining Kavalan people finally settling further south in Sinshe (Hsinshe, Xinshe, 新社) Village, a little village facing the Pacific Ocean, with Ocean Mountain Range at their back (See Picture 1). Others chose to settle in villages even further south along the Pacific coast, principally in Jangyuan, Taitung (台東樟原). Map 1 shows the two migration routes of the Kavalan people.



Sinshe Village--a sparsely populated village facing the Pacific Ocean on the southeast coast

<sup>3</sup> The Romanizations of the place names mentioned in this paper are conventional English translations adopted by each local government.

<sup>4</sup> Taiwan was at that time governed by the Qing Dynasty, but that was in name only. Many indigenous tribes were independent entities and did not come under the jurisdiction of the Qing Empire. For example, it was not until 1810 that a government office was set up to try to rule over the 36 Kavalan tribes.

The population of the Kavalan people has changed drastically over the last three hundred years. As shown in Table 1, the Kavalan population has over the centuries steadily declined in their homeland, Ilan, with just four Kavalans living there now, according to the official census of the Council of Indigenous Peoples, Executive Yuan.

TABLE 1. Population Change of the Kavalan in Ilan County

Year	1650 *	1852 *	1896*	1935 *	1969*	2007**
Population	9770	5507	2780	1544	app. 800	4

\* Data taken from Huang & Chang (1995:2)

\*\* Data taken from the Council of Indigenous Peoples, Executive Yuan (02/2007)



MAP 1. Migration Routes of the Kavalan people

Sinicization and the eventual loss of identity have combined to produce a sharp population decline in the two other counties where the Kavalan now live, i.e., Hualien and Taitung. As shown in Table 2, there are now just 650 Kavalan people residing in these two counties, 568 in Hualien and 82 in Taitung. In other words, 87% of the Kavalan people now live in Hualien County, with most of them concentrated in Fengbin Township (豐濱鄉).

TABLE 2. Population Change of the Kavalan in Hualien and Taitung

Year	1897*	1966*	2007**
Population	About 1000	1289	650

\* Data taken from Huang & Chang (1995:2)

\*\* Data taken from the Council of Indigenous Peoples, Executive Yuan (02/2007)

There are five villages in Fengbin Township: Gangkou (港口村), Sinshe (新社村), Jingpu (靜浦村), Jici (磯崎村), and Fengbin (豐濱村). As shown in Table 3, Sinshe (新社村) and Fengbin (豐濱村) villages are the only two significant settlements for the Kavalan. However, even in these two villages, the Kavalan are vastly outnumbered by the Amis (14 % vs. 86%), as shown in Table 4, which accounts in part for the decline of the state of health of the Kavalan language.

TABLE 3. Kavalan Population in Fengbin Township (豐濱鄉) (04/2007)

Village	Male	Female	Total
Gangkou港口	7	12	19
Sinshe新社	105	86	191
Jingpu靜浦	0	0	0
Jici磯崎	6	5	11
Fengbin豐濱	65	63	128
<b>Total</b>	<b>183</b>	<b>166</b>	<b>349</b>

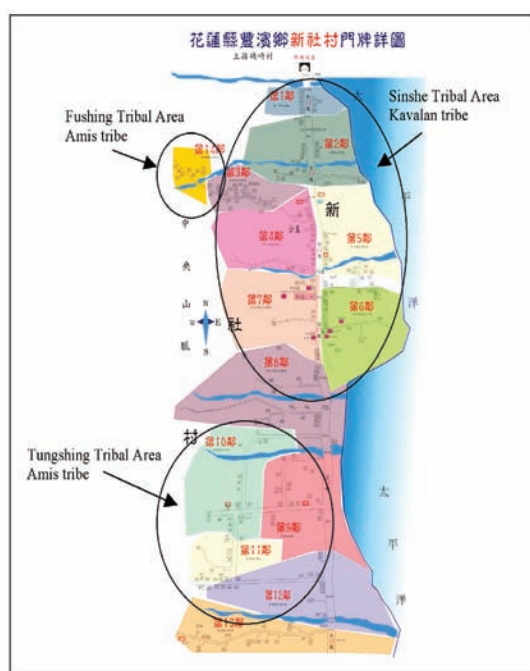
TABLE 4. Kavalan and Amis populations in Fengbin Township (04/2007)

Village Population	Male		Female		Total	
	Kavalan	Amis	Kavalan	Amis	Kavalan	Amis
Gangkou港口	7	437	12	334	19	771
Sinshe新社	105	249	86	211	191	460
Jingpu靜浦	0	442	0	296		738
Jici磯崎	6	98	5	110	11	208
Fengbin豐濱	65	853	63	681	128	1534
<b>Total</b>	183	2079	166	1632	349	3711

## 2.2 THE STATE OF HEALTH OF THE KAVALAN LANGUAGE

**2.2.1 BACKGROUND.** An early, and in fact the only sociolinguistic study on the Kavalan language was conducted more than a decade ago (Huang and Chang 1995). To update on the current state of health of the language, we did a village-wide sociolinguistic survey in Sinshe Village in April, 2007.

As shown in Map 2, the Kavalan people spread over eight neighborhood areas (from the First Neighborhood to the Eighth Neighborhood), which are collectively called Sinshe Tribal Area (新社部落). The Fourteenth Neighborhood is the Fushing Tribal Area (復興部落), where all the residents are Amis speakers. Another Amis tribal area, called Tungshing Tribal Area (東興部落), covers the Ninth to Thirteenth Neighborhoods.



MAP 2. A Detailed Map of Sinshe Village

(taken from the website of the Administration of Residents and Residence of Fengbin Township, Hualien, at <http://www.hl.gov.tw/977/upload/upload//index02.htm>)

**2.2.2 METHODOLOGY.** As mentioned earlier, the Kavalan population has over the centuries steadily declined. In Sinshe Village there were just 191 people in the official household registration record, which means the actual number of residents should be far less, since many people work and live in the cities, but still retain their names in the household registration. The Village is characterized by a high proportion of inter-marriages and a low proportion of the younger generation. Consequently, the respondents in this survey were mostly senior citizens. The criterion we used to select our respondents was that they have at least one Kavalan parent. There were in this survey a total of 12 female and 11 male

respondents, with a mean age of 63 and an average age of 62.87; the oldest participant was 87 and the youngest 42. Nearly all of our respondents were multi-lingual in Kavalan, Amis, Mandarin, and Taiwanese, and a few even spoke a fifth language, Sakizaya.<sup>5</sup> All of them were married, and had children.

In this sociolinguistic survey, we focus our attention on the language ability of the Kavalan people in the Sinshe Village. Language ability in this study is defined as “being able to use the language to communicate with the family members or with the community members.” In other words, we evaluate a speaker’s language ability in terms of whether s/he can use the language to communicate rather than whether s/he can spell a word correctly. Therefore, the questions we asked our respondents were three. (1) “What language or languages do you use as a means of communication when you talk to your parents (last generation) and when you talk to your children (next generation)?” (2) “In what language or languages do your parents talk to you?” (3) “In what language or languages do your children talk to you?”

We interviewed our respondents in their homes or at a place where they usually gather for social occasions, such as churches, grocery stores or weaving classroom. At each interview session, we first explained our intention to our respondents that we would like to know how and when they used Kavalan to communicate with other villagers. Then, we recorded each respondent’s bio-data, i.e., age, marital status, number of children, the tribes that his/her parents and grandparents belong to, and the tribes that his/her spouse belongs to. Next, we asked the three questions listed above to determine their language ability. For each language the respondent mentioned, we marked one point in that language column. Since all the respondents in this study were multilingual, their answers may not be limited to a single language.

**2.2.3 RESULTS AND DISCUSSIONS.** We list our findings in Table 5 below. In Table 5, the abbreviations K, A, T, M, and O stand for the Kavalan, Amis, Taiwanese, Mandarin and other languages, such as Sakizaya, Japanese or Hakka, respectively.

As shown in Table 5, all the respondents, i.e., 23/23, used Kavalan at home. 78.26% of their parents (i.e., 36 of the 46) used Kavalan, but 86.96% of their parents (i.e., 40/46) spoke Amis at home; since some of their parents were Amis. 73.91% of their children now in their 30s-40s also used Kavalan without difficulty. All of the respondents’ next generation also spoke Mandarin. Note the comparatively lower percentages among the respondents and their last generation on this measure.

In the home domain, all our senior respondents reported that they spoke Kavalan to the family members. However, if they talked to the younger family members, those below 20, they talked in Kavalan and the younger family members replied in Mandarin or Taiwanese.

---

<sup>5</sup> Although Kavalan is still actively used in the village, outside the home domains the Kavalans speak different languages to different co-participants, as expected. When they go grocery shopping in a Taiwanese’s store, they speak Taiwanese. When they meet their Amis neighbors, they chat in Amis. Although Kavalan and Amis live in different parts of the Village, as shown in Map 2 above, they go to the same churches, Catholic or Christian, there being only one each in Sinshe. In the churches, especially the Catholic, the Kavalans tend to switch to Amis, since the priest is an Amis, and the Bible is in Amis. At present there is no Kavalan Bible. All these suggest that Kavalan is in some sense a minority language even in Sinshe Village.

TABLE 5. Language Ability of the Kavalan people in Sinshe Village

Language	Last Generation					Respondents					Next Generation				
	K	A	T	M	O	K	A	T	M	O	K	A	T	M	O
	36/46	40/46	23/46	16/46	23/46	23/23	21/23	14/23	16/23	9/23	17/23	9/23	14/23	23/23	5/23
Percentage	78.26%	86.96%	50%	34.78%	50%	100%	91.3%	60.87%	69.57%	39.13%	73.91%	39.13%	60.87%	100%	21.74%

K: Kavalan; A: Amis; T: Taiwanese; M: Mandarin; O: Others (esp. Sakizaya, Japanese, and Hakka)

TABLE 6. Language Abilities of the Kavalan (comparative)

Language	Last Generation					Respondents					Next Generation				
	K	A	T	M	O	K	A	T	M	O	K	A	T	M	O
	24/32	27/32	12/32	8/32	11/32	16/16	15/16	8/16	9/16	8/16	15/16	8/16	13/16	16/16	5/16
51 ↑ *	75%	84.38%	37.5%	25%	34.38%	100%	93.75%	50%	56.25%	50%	93.75%	50%	81.25%	100%	31.25%
	12/14	13/14	11/14	8/14	12/14	7/7	6/7	6/7	7/7	1/7	2/7	1/7	1/7	7/7	0/7
50 ↓ **	85.71%	92.86%	78.57%	57.12%	85.71%	100%	85.71%	85.71%	100%	14.29%	28.57%	14.29%	14.29%	100%	0%

\* Respondents who are older than 51 years old.

\*\* Respondents who are younger than 50 years old.

All of the respondents reported that almost all the children younger than 20 could hardly understand Kavalan, let alone speak fluent Kavalan. Consequently, the seniors often had to switch to Mandarin or Taiwanese to help the process along. The situation is predictably considerably worse for families residing in cities where environmental support for the use of Kavalan is non-existent (Huang and Chang 1995).<sup>6</sup>

We tried to find some young respondents in their 20s in Sinshe, but to no avail. People in their 20s-40s tend to live and work in the cities, since there are simply no jobs available in the village. Those still staying in the village are those older than 50, or those younger than 15, youngsters who are still in school.

There were a total of seven respondents younger than 50 in our survey, which means their children might be in their 20s. We thus separated these respondents from the others, and we arrived at Table 6.

As shown in Table 6, almost all of the children, about 94%, of the respondents older than 51 speak Kavalan, while only less than 30% of the children of the interviewees younger than 50 do. Given the limitedness of our sample size, the true picture of the rate of language transmission among these children is probably somewhere between these two extremes.

At this point, it may be of some interest to compare our findings with those in Huang and Chang's (1995), as shown in Table 7.

TABLE 7. Huang and Chang's Survey (1995:8)

Language		Last Generation			Respondents			Next Generation		
		K	A	T	K	A	T	K	A	T
Sinshe		23/28	16/28	23/28	13/14	13/14	12/14	<b>12/14</b>	6/14	14/14
	Percentage	82.1%	57.1%	82.1%	93%	93%	85.7%	<b>85.7%</b>	46.1%	100%
Taipei		64/64	63/64	54/64	31/32	28/32	32/32	<b>3/27</b>	2/27	21/27
	Percentage	100%	98.4%	84.4%	96.8%	87.5%	100%	<b>11.1%</b>	7.4%	77.7%

In their study, Huang and Chang conducted two surveys, one in Sinshe and one in Taipei.<sup>7</sup> The average age of Huang and Chang's (ibid: 5) Taipei respondents was 36.5, and that of their Sinshe respondents was 58.5; therefore, the average age of the next generation of the Taipei respondents was estimated to be 17 or younger, and that of the Sinshe respondents was about 30 or older. Huang and Chang's (ibid) survey shows that the language transmission rate in Sinshe stood at 86%, while that in Taipei was a low of 12%. Their finding is quite similar to ours, since our survey shows that the older respondents were able to transmit their mother language to their next generation (about 94%), while the younger respondents can barely do so (only about 29%).

Nonetheless, there are two exceptional cases in our survey. One case involves two

<sup>6</sup> The environmental support here means the family members, friends, the community or even the school and the work places.

<sup>7</sup> Most of the Kavalan migrating into the Taipei area reside in Banqiao (板橋) and Shulin (樹林) areas.



little children, a four-year-old boy and his 2-year-old younger brother, who live with their Kavalan-speaking grandmother in Sinshe, and are acquiring the language as their first language, as their parents, the Kavalan father married to a Vietnamese bride, have to work in Taipei and do not have time to take care of them. In the other case, two teens are living with their parents and can speak Kavalan because their father, one of our younger respondents, 42, deliberately “creates the home environment” for them; that is, their ‘language crisis-conscious’ father asks them to speak Kavalan in the home.

As suggested above, another settlement area for the Kavalan people is Jiali Village of Sincheng Township in Hualien County (花蓮縣新城鄉), the midway homeland to many of the ancestors of those now living in Sinshe. As shown in Table 8, there are now only 105 Kavalans living in the town of Sincheng, 65 of whom live in Jiali Village (嘉里村), where the Kaliwan Incident took place and was once known as Kaliwan Village.

TABLE 8. The Kavalan Population in Sincheng Township, Hualien (03/2007)

Tribe Village	Amis	Kavalan
Dahan (大漢村)	983	5
Beipu (北埔村)	772	15
Jialin (佳林村)	146	2
Kangle (康樂村)	398	1
Shunan (順安村)	67	0
Sincheng (新城村)	81	3
Jiali (嘉里村)	492	<b>65</b>
Jiasin (嘉新村)	336	14
<b>Total</b>	<b>3275</b>	<b>105</b>

We were able to locate and interview two Kavalan seniors in the Village. One gentleman was Tuyaw (李抵搖), 88, who is married to a Sakizaya from Yuli (玉里) and speaks Kavalan, Sakizaya, Japanese, Amis and Taiwanese. Tuyaw remembered that his grandfather came from Sanshing, Ilan (宜蘭三星), and that his grandparents and parents could speak fluent Kavalan, Amis and Sakizaya. He could not recall the approximate time period when everyone in the neighborhood started to discontinue the use of Kavalan altogether. The most common languages he uses in his daily routines are Taiwanese, Amis, and, occasionally, Japanese. He hardly ever has any opportunity to speak Kavalan now, since no one speaks the language in the neighborhood. None of his seven children speaks either Sakizaya or Amis, much less Kavalan, the minority among the minority languages in the town.

Another respondent was the Kavalan wife of a former *Tapang no yaro* (chief of the Local Amis, 阿美族頭目) Jiang Jia-zou (江加走). Mrs. Jiang, 71, speaks fluent Taiwanese like a native Taiwanese speaker, and does not keep a Kavalan name. She spoke Kavalan while she was little, but, like Tuyaw, has hardly ever used it since then, since her neigh-

bors, relatives, or even her four children do not speak the language. While Kavalan is thriving forty miles down south in Sinshe, it is reasonable to suggest that Kavalan was already languishing in Sincheng when Mrs. Jiang was young and soon became extinct in the neighborhood as it lost its status even as a home language or a language of local social communication.

Huang and Chang (1995:7) explained that the rapid decline and eventual loss of the Kavalan language in areas outside of Sinshe Village is triggered by a combination of factors, namely, a high percentage of intermarriages, the ‘destabilizing effect’ of the more prestigious languages, i.e., Mandarin, Taiwanese and Amis, used in the wider communities, and, concomitantly, the relatively low degree of self-identity. These elements are still exerting their powerful forces. To be sure, indigenous issues have had a high profile over the past decade and the government has encouraged their rights to self-identity and the resumption of indigenous names and a number of policy changes in the direction of greater respect for the indigenous languages and heritage. ‘The indigenous language nest program’ has also been in effect for a number of years now--- these changes mean removal of social and political forces that have helped shape the language inequality in the first place and may yet provide structural conditions for the preservation, or at least slow down the process of decline, of indigenous languages.

**2.3 PROPOSED STRATEGIES FOR REVITALIZING THE KAVALAN LANGUAGE.** As part of a new educational policy, indigenous languages have been taught in the elementary schools since September 2001. Although these language lessons typically run only 1 or 2 hours a week, they are a step forward toward preserving the indigenous languages (cf. Lillian Huang 2007). Nonetheless, the current state of the Kavalan language teaching in the Village does not bode well for Kavalan. In Sinshe Village, there is only one elementary school, and it teaches both Amis and Kavalan. The students taking Amis far outnumber those taking Kavalan, since the Amis population is the majority in the village. The teacher teaching Amis is an Amis, who is also a regular member of the school faculty, while the Kavalan teachers are not. When the students go on to junior high school, in the neighboring village, only Amis but not Kavalan is taught. Furthermore, there is no other way for these Kavalan children to continue their education in Kavalan, except from their families.

Although the Council of Indigenous Peoples (CIP) has launched a six-year program since 2003 to help revitalize indigenous languages in both rural and urban areas (cf. Lillian Huang 2007), school or the community leaders in Sinshe Village have failed to propose any language revitalizing programs with which to apply for grants from the CIP, thus forestalling Kavalan language revitalizing efforts in the village, since without government grants, they could not set up community-based language classrooms to train language teachers.

Concerned linguists and community leaders must join hands to help come up with language revitalizing programs and apply for funding from the CIP (i) to set up public Kavalan classrooms, and (ii) to train more Kavalan teachers committed to Kavalan language teaching. To be sure, if a language is not spoken in the home, classroom teaching might seem to be a superficial and cosmetic measure, but it can have other positive functions. Many Kavalan parents expect their children to get good grades in school in order to go to a better high school and college, and, eventually, get a better job. Classroom teaching can

thus make learning Kavalan matter in this regard. Moreover, classroom teaching is also a direct way for the Kavalan children to know about their own language, culture and history. Armed with this knowledge they can then go on to document their legends, folklores and songs and even work on their own bible. Although there are now a number of studies related to the Kavalan languages (Li 2007), including a recently published Kavalan Dictionary (Li and Tsuchida 2006), no Kavalan version of the Bible is yet available, and so the Kavalan churchgoers in Sinshe have to use the Amis version, to the detriment of the Kavalan language in the community. Compiling a Kavalan version of the Bible is sorely needed.

In the next section, we will introduce the NTU Corpus of Formosan Languages, which is built with an attempt to document some of the most seriously endangered Formosan Languages.

**3. NTU CORPUS OF FORMOSAN LANGUAGES.** Digital Archiving of the Yami Language at Providence University (Rau and Yang 2007; Rau, Yang and Dong 2007) stands as the first attempt to provide public access to the language. In addition, Academia Sinica's Formosan Language Archive (Zeitoun et al. 2003; Zeitoun and Yu 2005) appears to be a large-scale digital archive with an attempt to document all the Formosan languages, including their dialects.<sup>8</sup>

The NTU (National Taiwan University) Corpus of Formosan Languages demonstrates our attempt not only to document some seriously endangered Formosan languages, but to provide further public and user-friendly access to both specialists and non-specialists.<sup>9</sup> There are two special features of the NTU Corpus, i.e., dictionary and search, which we discuss in detail below. Moreover, the NTU Corpus of Formosan Languages is the first corpus to document spoken data in terms of prosodic units, i.e., the Intonation Unit (IU), which is defined as a stretch of discourse falling under a single coherent intonation contour (Chafe 1987, 1993, 1994; Du Bois et al. 1993; Schuetze-Coburn 1993; Schuetze-Coburn et al. 1991; Tao 1993). Although natural spoken language is often found to have a high proportion of pauses, hesitations, fillers, repetitions, and false starts, they are important for us to learn more about the pragmatic and cultural aspects of a spoken language. A detailed study of fragments of conversation reveals that fillers, repairs and repetitions are important interactional strategies used by the speech participants to hold the floor, to plan for language production, to do lexical searching, and so on. Fillers, repairs and repetitions are thus essential in that they enable the conversation to go on without much difficulty by sending out signals of the speaker's next move and intention (Schegloff 1980, 1988, 1991, 1996; Sack, Schegloff and Jefferson 1974; Huang 1993, 1999).

**3.1 AN OVERVIEW OF NTU CORPUS OF FORMOSAN LANGUAGES.** The main purpose of the NTU Corpus of Formosan Languages is to document some of the endangered languages spoken in Taiwan, such as Kavalan, Saisiyat and Tsou. It was part of the projects of the Multimedia Laboratory operated by the Center for Information and

---

<sup>8</sup> At present, Academia Sinica's digital archiving is still under construction; many language corpora are listed but contain no text.

<sup>9</sup> See Sung et al. (submitted) for a detailed description of the NTU Corpus of Formosan Languages.

Electronics Technologies at National Taiwan University,<sup>10</sup> with an attempt to establish a standard for the creation of linguistic databases through the application of information technology to linguistics research.

The corpus contains face-to-face conversations and narratives, including both natural and elicited narratives for cross-linguistic research. The materials for elicited narratives are based on the Pear Story (Chafe 1980) and Frog Story (Mayer 1980).<sup>11</sup> The NTU corpus is composed of spoken texts in Saisiyat, Kavalan, Amis and Tsou.<sup>12</sup> At present, a small corpus of Kavalan with just four narrative texts has been placed online, which runs to about ten minutes, for a total of 228 IUs, though we have collected and transcribed a total of 21 Kavalan texts, including 15 narratives and 6 conversations, which together run to about 136 minutes. The narrative texts are 11 elicited narratives, including four pear stories and seven frog stories, and 4 other narratives. The conversations cover six face-to-face conversations between acquaintances or relatives, which together run to about 61 minutes. All these texts will be put on line in the near future.

**3.2 TRANSCRIBING SPOKEN DATA.** The process of transcribing spoken data is both tedious and time-consuming.<sup>13</sup> We first sound-record or/and video-tape spoken texts. After a spoken text is collected, our graduate assistants then help transcribe the raw data, and tag and annotate the transcribed texts according to a prepared coding list. Then other assistants double check the annotated text.

Since our corpus data are natural spoken narratives and conversations, in order to reflect and record the discourse information, such as pauses, false starts, repairs, and so on, the transcription needs to meet not only grammatical but also discourse coding standards. Our transcription of the discourse information mostly follows Du Bois et al. (1993), a *de facto* standard in the linguistic community. After the tagging, annotating and double-checking, the transcribed texts are submitted to the corpus programmer before they are finally put on line. The whole transcription process can be diagrammed as in Figure 1.

The texts in our system are stored in Unicode (UTF-8 encoding); the advantage of using such an encoding form is that it is easy to incorporate other languages into our an-

---

<sup>10</sup> They are Graduate Institute of Linguistics, Department of Information Management, Department of Library and Information Science, Department of Computer Science and Information Engineering, Department of Electrical Engineering, Department of Journalism, and Department of Drama and Theater) and colleges (College of Electrical Engineering and Computer Science, College of Liberal Arts, College of Social Sciences, and College of Management)

<sup>11</sup> Frog stories are elicited narratives: our informants are asked to tell a story while looking at the pictorial book *Frog, Where Are You?* (Mayer 1969, reprinted in 1980). This pictorial wordless book tells the adventure story of a little boy and his dog searching for his frog that got out into the woods.

<sup>12</sup> The corpus of Saisiyat has 22 texts, including three conversations, five Pear stories, eight Frog stories, four Saisiyat legends and two daily life narratives, which together run to 118 minutes, for a total of 3437 IUs. The Corpora of Amis and Tsou each contain two narratives, one Frog story and one Pear story, which run to five minutes with a total of 138 IUs and eight minutes with a total of 237 IUs, respectively.

<sup>13</sup> According to our rough estimation, one-minute of raw data requires ten to twelve hours of working time.

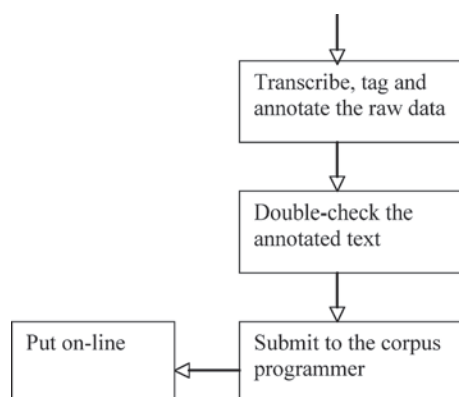


FIGURE 1. Procedure of data transcription and submission

notation system which adopt other writing or phonetic systems, such as IPA. If some of the tribes decide to adopt non-ASCII letters, such as “ê Â ú Ò Ä”, into their writing systems, our programs can process them correctly with no need of modification or transformation.

**3.3 ACCESSIBILITY AND SPECIAL FEATURES.** There are two special features of the NTU Corpus, i.e., dictionary and search, which we will illustrate with examples below. One special feature is that our system can automatically generate an online dictionary, with the information of the total number of word tokens. The count of tokens is updated as new texts are uploaded. As shown in Figure 2, when choosing a language and entering the corpus, the user can find a “Dump dictionary” function on the top of the list of the texts. The dictionary can be printed out at a marginal cost, and it can also be cut and pasted for any linguistic analysis.

Another special feature of our corpus is that it allows users to search for any specific target word or morpheme in English, Chinese or any Formosan languages. For example, if a user wants to know how to say the English word *know* in these Formosan languages, he may type in the English word ‘know’ and then he may find that in Kavalan, the equivalent word is *supaR*, while in Saisiyat, there are two words equivalent to the English word *know*, *sekela* and *ra:am*. In Amis, the equivalent is *ma-fana*; and in Tsou, it is *cohivi*.

In the Search function page, one can also search for a particular lexeme in a corpus; for example, one may want to search for the distribution of the lexeme *Rayngu* ‘not know; not able to’ in Kavalan. He may select the language Kavalan, and type in the key word, *Rayngu*; then, he can find all the related data, as shown in Figure 3.

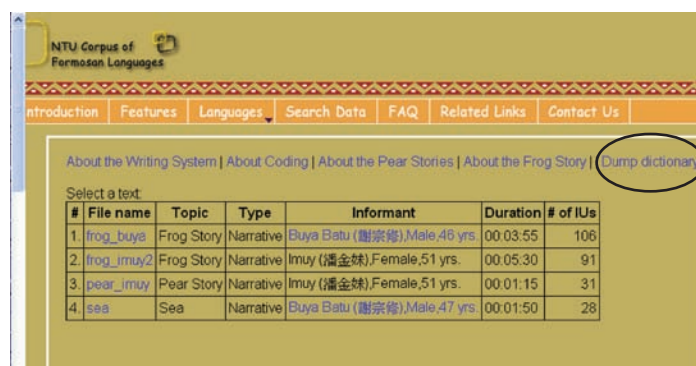
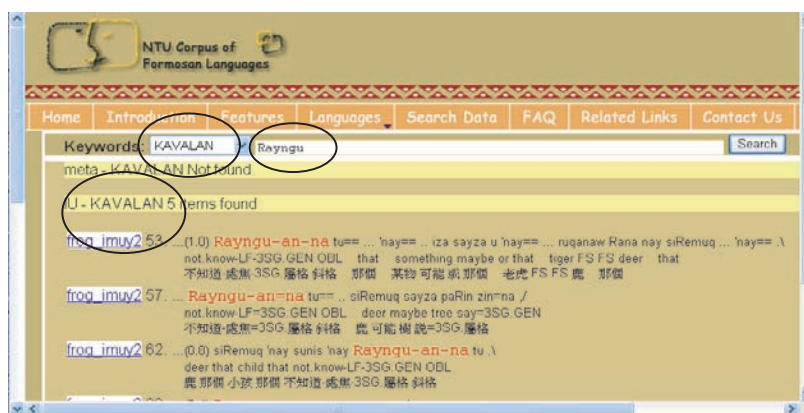


FIGURE 2. The function [Dump dictionary] in the Kavalan Corpus.

FIGURE 3. Pop-up search result of the lexeme *Rayngu* in the Kavalan corpus

Moreover, users can search for a particular lexeme across the languages in the corpora, e.g., *ma-* in Kavalan, Amis, and Saisiyat. Since each text and each Formosa language in our corpus is stored in a cross-related file with the same normalized tables, cross-text and cross-language search can be executed in a single command (cf. Sung et al., submitted). The number of tokens of the searched lexeme is shown at the same time. For example, if a user wants to investigate the syntactic behaviour of the marker *ma-* in Formosan languages (say, Kavalan, Saisiyat, Amis and so on), he/she can type in the key morpheme *ma-* in the search page, wait for a second and then all the related data in different texts and (Formosan) languages will come out, as shown below in Figure 4.



FIGURE 4. Pop-up search results of the prefix *ma-* (Page 1 of 31 pages)

With the rapid progress in internet technology and the processing techniques of natural linguistic database, the creation of a language database has become a most effective means of recording and preserving precious linguistic data. The NTU Corpus is structured in a way that enables any user, linguist or not, who is interested in Austronesian languages and culture, to gain access to the rich and valuable linguistic data available through a diverse array of format in the most convenient means.

**5. CONCLUSION.** In the preceding sections we have provided a fairly detailed survey of the state of health of the Kavalan language. We have also looked at the workings of the NTU Corpus of Formosan Languages, which is built with an attempt to document some of the endangered languages spoken in Taiwan, languages that are Taiwan's gifts to the world (Diamond 2000). Documenting Kavalan as well as other Formosan languages is an ongoing project for us. Budgetary constraints and shortage of staff have meant that we have not been able to proceed at a pace we would have liked it to be. When we are done, however, we hope the NTU Corpus of Formosan Languages will ultimately prove to be a valuable research tool to the academic community.

## REFERENCES

- BRENZINGER, MATTHIAS, ed. 1992. *Language death: Factual and theoretical explorations with special reference to East Africa*. Berlin: Mouton de Gruyter.
- CHAFE, WALLACE L. 1980. The deployment of consciousness in the production of a narrative. In *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, ed. by Wallace Chafe, 9-50. Norwood, N.J.: Ablex Publishing Corp.
- CHAFE, WALLACE L. 1987. Cognitive constraints on information flow. In *Coherence and grounding in discourse*, ed. by Russell S. Tomlin, 21-51. Norwood, N.J.: Ablex.
- CHAFE, WALLACE L. 1993. Prosodic and functional units of language. In *Talking data: Transcription and coding in discourse research*, ed. by Jane A. Edwards and Martin D. Lampert, 33-43. Hillsdale, N.J.: L. Erlbaum.
- CHAFE, WALLACE L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- CRAWFORD, JAMES. 1995. Endangered native American languages: What is to be done, and why? *The Bilingual Research Journal* 19(1):17-38.
- DIAMOND, JARED M. 2000. Linguistics: Taiwan's gift to the world. *Nature* 403:709-10.
- DU BOIS, J. W, STEPHAN SCHUETZE-COBURN, SUSANNA CUMMING, and DANAE PAOLINO. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding for language research*, ed. by Jane A. Edwards and Martin D. Lampert, 45-90. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-95.
- HUANG, LILLIAN M. 2007. Strategies in revitalizing indigenous languages in Taiwan. Pre-Conference Proceedings of the International Conference on Austronesian Endangered Language Documentation, 25-45. Taichung: Providence University.
- HUANG, SHUANFAN. 1993. Pause as a window on the mind and the grammar—evidence from spoken Chinese discourse. Paper presented at the Workshop on Interfaces and the Chinese Language, June 30-August 6, 1993, in Ohio State University.
- HUANG, SHUANFAN. 1999. The emergence of a grammatical category definite article in spoken Chinese. *Journal of Pragmatics* 31(1):77-94.
- HUANG, SHUANFAN, and H. 1995. Kavalan: A sociolinguistic study. *Ilan Wenxian (Ilan Journal of History)* 14:1-13. (In Chinese)
- KRAUSS, MICHAEL. 1992. The world's languages in crisis. *Language* 68:6-10.
- LEHMANN, CHRISTIAN. 2001. Language documentation: A program. In *Aspects of typology and universals*, ed. by Walter Bisang, 83-97. Berlin: Akademie Verlag.
- LI, PAUL JEN-KUEI. 2007. Documentation of the most endangered Formosan languages. Pre-Conference Proceedings of the International Conference on Austronesian Endangered Language Documentation, 1-12. Taichung: Providence University.
- LI, PAUL JEN-KUEI, and SHIGERU TSUCHIDA. 2006. *Kavalan Dictionary*. Taipei: Academia Sinica.
- MAYER, MERCER. 1980. *Frog, where are you?* N.Y.: Dial Books.
- RAU, D. VICTORIA, and MENG-CHIEN YANG. 2007. e-Learning in endangered language documentation and revitalization. Pre-Conference Proceedings of the International Conference on Austronesian Endangered Language Documentation, 101-22. Taichung: Providence University.
- RAU, D. VICTORIA, MENG-CHIEN YANG, and MAA-NEU DONG. 2007. Endangered language documentation and transmission. *Journal of National Council of Less Commonly Taught Languages (NCOLCTL)*. University of Wisconsin at Madison. 53-76.



- ROBINS, ROBERT. H., and EUGENIUS M. UHLENBECK, eds. 1991. *Endangered languages*. Oxford: Berg.
- SACK, HARVEY, EMANUEL A. SCHEGLOFF and GAIL JEFFERSON. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696-735.
- SCHEGLOFF, EMANUEL A. 1980. Preliminaries to preliminaries: "Can I ask you a question?" *Sociological Inquiry* 50:104-51.
- SCHEGLOFF, EMANUEL A. 1988. Pre-sequences and indirection. *Journal of Pragmatics* 12:55-62.
- SCHEGLOFF, EMANUEL A. 1991. Conversation analysis and socially share cognition. In *Perspectives on socially shared cognition*, ed. by Lauren B. Resnick, John M. Levine and Stephanie D. Teasley, 150-71. Washington D.C.: American Psychological Association.
- SCHEGLOFF, EMANUEL A. 1996. Turn organization. In *Interaction and grammar*, ed. by Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson, 52-133. Cambridge University Press.
- SCHMIDT, ANNETTE. 1990. *The loss of Australia's aboriginal language heritage*. Canberra: Aboriginal Studies Press.
- SCHUETZE-COBURN, STEPHAN. 1993. *Prosody, grammar, and discourse pragmatics: Organizational principles of information flow in German conversational narratives*. PhD diss., University of California at Los Angeles.
- SCHUETZE-COBURN, STEPHAN, MARIAN SHAPLEY and ELIZABETH WEBER. 1991. Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech* 34(2):207-34.
- SUNG, LI-MAY, LILY I-WEN SU, FUHUI HSIEH and ZHEMIN LIN. Submitted. Design of a multimedia corpus of Formosan languages.
- TAO, HONGYIN. 1996. *Units in Mandarin conversation: Prosody, discourse, and grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- ZEITOUN, ELIZABETH, CHING-HUA YU, and CUI-XIA WENG. 2003. The Formosan language archive: Development of a multimedia tool to salvage the languages and oral traditions of the indigenous tribes of Taiwan. *Oceanic Linguistics* 42(1):218-32.
- ZEITOUN, ELIZABETH, and CHING-HUA YU. 2005. The Formosan language archive: Linguistic analysis and language processing. *Computational Linguistics and Chinese Language Processing* 10(2):167-200

Fuhui Hsieh  
hsiehfh@ttu.edu.tw

Shuanfan Huang  
sfhuang@ntu.edu.tw

## E-Learning in Endangered Language Documentation and Revitalization

D. Victoria Rau and Meng-Chien Yang

*Providence University, Taiwan*

This paper analyzes the application of e-learning in the revitalization of endangered languages. It outlines the areas in which e-learning is efficacious, the attitudes of the indigenous language teachers to e-learning, the feelings of the Yami community toward this kind of pedagogy, and the reactions of the users, mostly young and adolescent learners of Yami. The findings are based on the results of surveys and in-depth studies in the Yami community and also on surveys made in a nation-wide seminar that enrolled teachers of the majority of the still-spoken aboriginal languages in Taiwan. Both qualitative and quantitative methods were used to gather empirical data to address questions in the following three areas: (1) the contexts of developing e-Learning materials for endangered indigenous languages in Taiwan, (2) the indigenous language teachers' perceptions of e-Learning in Taiwan, and (3) the attitudes of the Yami community on Orchid Island toward e-Learning. This paper provides a model for the many language revitalization projects underway in Taiwan and worldwide to take advantage of e-Learning. It also provides guidelines that enable each project to better understand the kinds of e-Learning that work best and the type of community outreach necessary to make e-Learning acceptable and efficacious.

**1. INTRODUCTION.** Language documentation has long been recognized as an effective means of language preservation. However, very little effort has been made to explore how information and communication technology (ICT) can be used in the promotion of minority language teaching and research in connection with language documentation.

In our previous studies, a conceptual framework has been developed for integrating e-Learning (computer-based instruction) into language documentation (Yang and Rau 2005). In this framework, as shown in Figure 1, language documentation and e-Learning programs are two separate modules. The language documentation module has two phases: preparation and digitalization. The e-Learning program module also has two phases: program development and program deployment. Each phase has two to three conceptual sub-phases representing the work to be completed. Below the two modules are the three groups of participants in this framework. Each group of participants has their own domain of knowledge and assigned tasks. We are using this framework to document and produce e-Learning materials for the Taiwanese aboriginal language Yami.

Rau and Yang (forthcoming 2008) have also identified three groups of participants and illustrated how they could collaborate to develop Yami e-Learning materials. The process of developing e-Learning materials for Yami is described in the framework of nexus analysis. The targeted learners are university students, focusing primarily on graduate students who are interested in Taiwanese local languages and cultures, with the potential of extension to Yami community members who are two or more generations removed.

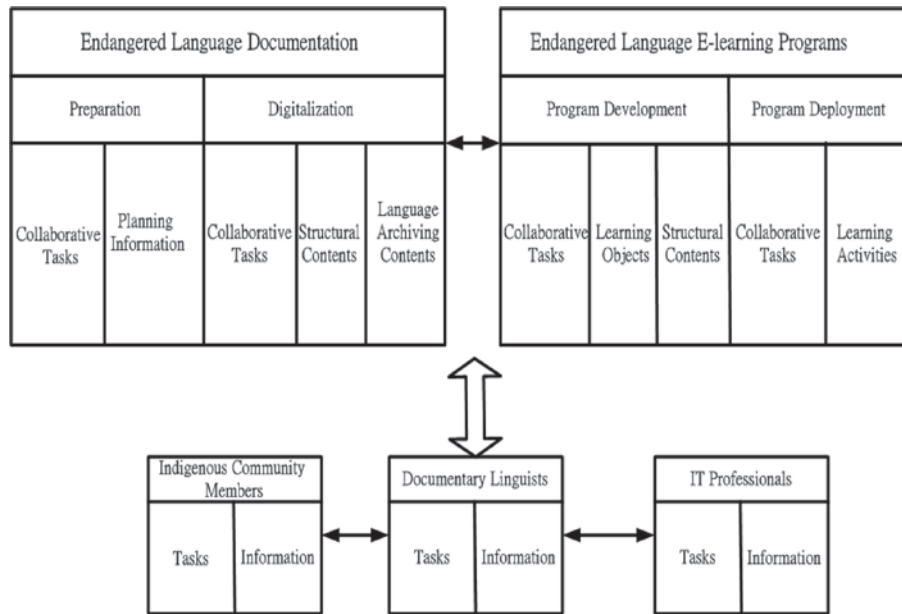


FIGURE 1. Framework for Endangered Language Documentation and e-Learning Development

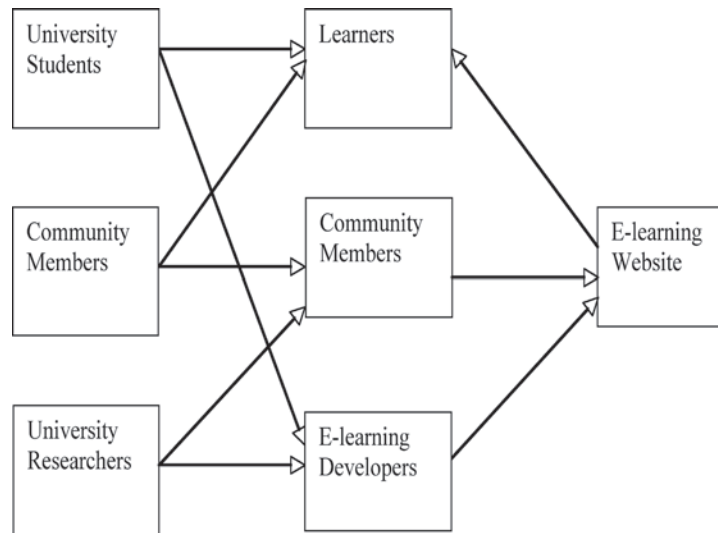


FIGURE 2. Participants in Yami e-Learning (adapted from Ward and van Genabith 2003)



FIGURE 3. The e-Learning platform for the Yami language

The community members were brought into the picture through many initiatives of the researchers to facilitate continued partnership of the community members with the research team.

The university researchers are primarily content providers and e-Learning developers, with the potential goal of becoming learners. The three major figures consist of an applied linguist who did extensive research on Yami, a computer scientist with an expertise in digital archiving and e-Learning, and a native Yami who had knowledge and experience in linguistic analysis and Yami language teaching.

As shown in Figure 2, each participant serves at least two roles in the process.

In rethinking pedagogical models for e-Learning, Rau, Yang and Dong (2007) provided a historical account of Yami language teaching, from a grammatical syllabus (Dong and Rau 2000) to a communicative syllabus using a multimedia CD as a resource (Rau et al. 2005), to the development of interactive on-line learning based on our digital archiving project, as illustrated in Figure 3.

The technology of e-Learning has dramatically changed the genre of the language textbook, the context of language teaching and learning, and the traditional roles of language teachers and learners. The key benefits of e-Learning, such as the ability to meet individual learning needs, access a wider range of resources, and be exposed to multimedia materials, including both images and sounds, are particularly useful for the design of teaching materials for less commonly taught languages and/or endangered languages, since such textbooks are not readily available and are extremely costly to produce. However, little research has been conducted to explore the potential benefits of e-Learning for the promotion of endangered language revitalization.

Some researchers seem to be optimistic and try to integrate traditional cultural patterns into the development of learning materials (e.g., Csató and Nathan 2003), whereas others tend to have reserved attitudes and reported resistance from the local communities (e.g., Villa 2002). Of course, these studies have different participants from different geographical areas and cultural backgrounds. The attitudes of the indigenous people toward the Internet and e-Learning technology are hard to generalize. Based on our framework, the attitude of the language teachers is critical when developing e-Learning programs for the endangered languages. When an e-Learning development project is initiated in an indigenous community, a mutual understanding should be established. During the development and deployment of the e-Learning program, a retrospective review regarding the effectiveness of the e-Learning program should be obtained. However, very little research has explored indigenous teachers' attitudes towards e-Learning and the results have not been very positive.

Eisenlohr (2004) reported that some researchers faced resistance from the local community, who questioned the appropriateness of creating and distributing digital archives of the endangered languages. The indigenous peoples may think of the creation and online publication of such an archive as a threat to their own language (Villa 2002). In addition, the fact that very few indigenous people have the capability to use and manage the computer software to document their own language is another major obstacle. Researchers could face the difficult situation of not being able to contribute to language revitalization at all after documenting an endangered language.

In contrast with other endangered language communities where a solid Internet infrastructure is usually lacking, Taiwan has the highest number of digital households in the world (PARKS 2005), followed by Korea and the USA, and is ranked thirteenth in the world in individual, business, and government uses of information communication and technology (WEF 2007). In 2002, the Taiwanese government also launched 77 national e-Learning plans to promote e-Learning environments for academic and scientific research, to narrow the gap of the digital divide (less Internet access east of the central mountain range), and to advance the development of mobile learning and industrial markets (Liou 2005).

In indigenous communities in Taiwan, students can access the Internet and surf the web at school. However, the indigenous language speakers and language teachers over 50 years of age have limited knowledge of the Internet and computers. Thus the primary question becomes whether they are willing to use the computer and e-Learning as a tool for language revitalization?

The major focus of this paper is to provide a summative assessment of the impact of our e-Learning materials, funded by the Hans Rausing Endangered Languages Project, for the purpose of endangered language documentation and revitalization. The research questions for this study are as follows:

1. What are the contexts of developing e-Learning materials for endangered indigenous languages in Taiwan?
2. What are the indigenous language teachers' perceptions of e-Learning in Taiwan?
3. What are the attitudes of the Yami community on Orchid Island toward e-Learning?

Both qualitative and quantitative methods were used to present empirical data and interpretations of our assessment of e-Learning. We begin with a general introduction to the Yami community and their languages, followed by a critical evaluation of the continua of biliteracy (Hornberger 2003) in Taiwan as the background for language documentation and e-Learning for endangered languages. The second half of the paper presents two studies on evaluation of our proposed framework and the e-Learning program. The first study is a quantitative analysis of the indigenous teachers' perceptions of e-Learning in Taiwan. The second study is a description of the results of our interviews with Yami community members regarding their attitudes towards e-Learning.

**2. SOCIOLINGUISTIC BACKGROUND ON YAMI.** Yami is a Philippine language, spoken by approximately 3000 speakers on Orchid Island, located at the northern tip of the Batanes Province of the Philippines. Yami is a Batanic language, closely related to Itbayat, Ivatan, and Babuyan. Politically part of the territory of the Republic of China, Mandarin Chinese is the official language.

According to Rau (1995b), Yami constitute 93% of the 3007 residents on Orchid Island. Almost half of the population on the island is either above 50 or below 20 years old. Young adults usually seek employment in Taiwan. Yami people above 60 years of age are mostly monolingual in Yami, whereas those below 20 consider Mandarin Chinese their L1 and Yami their L2 (Chen 1998). Young adults code-switch between Yami and Chinese in communication. Iraralay is the only community of the six villages on the island where children still use Yami for daily interaction (Lin 2007). Two Yami dialects, Iraralay and Imowrod are mutually intelligible with some lexical differences and systematic vowel changes (Rau, Chang and Dong, forthcoming). The more conservative Imowrod dialect is used as the basis for standard orthography.

Although Yami has been offered as an elective in elementary school since 1998, Yami is gradually being replaced by Mandarin Chinese in daily use. Among the junior high school students on Orchid Island, 60% either believed Yami would die eventually or were uncertain about the fate of the language (Rau, 1995b)

Adult speakers use both Chinese and Yami in daily communication. The medium of education in schools is exclusively Chinese. Yami is used primarily in Christian church services and traditional ceremonies. Translation of the New Testament of the Bible into Yami was completed in 1995. There is a locally run radio station on the island, managed by a Christian organization, broadcasting programs in Chinese and Yami.

Three different orthographies have been proposed, all based on Roman alphabets with minor differences in the representations of trill vs. flap and glide vs. vowel. One is used in the newly translated Bible, developed in collaboration between SIL missionaries and local pastors. The second was announced by the Council of Indigenous Peoples in 2005 in an effort to standardize the writing systems of all Austronesian languages in Taiwan. The third was jointly developed by Dong and Rau during their collaboration and used in their texts, teaching materials, and Yami dictionary (Rau and Dong, 2006). Except in the teaching of Yami language in primary and secondary school and teacher training workshops, Yami orthography is not in general use by anyone in the speech community. The community has various degrees of literacy in Chinese (95%) while the most educated (less than 1%) are also literate in English (Rau, 1995b).

**2.1. CONTINUA OF BILITERACY.** The factors influencing biliteracy in Taiwanese society, dominated by the Mandarin and Southern Min Chinese dialect speakers, are very different from those in the indigenous communities. In this section, we examine the power relations of the language communities in Taiwan, based on Hornberger and Skilton-Sylvester's (2003: 39) model. To introduce e-Learning into an endangered indigenous language community, we need to be aware of the context, development, content and media of biliteracy in the indigenous communities in Taiwan.

Context of biliteracy: Indigenous communities in Taiwan are oral monolingual in their ethnic tradition but monoliterate in Mandarin Chinese.

Development of biliteracy: Development of indigenous language literacy is considered second language acquisition for teenagers living in their own communities and foreign language acquisition for those living in metropolitan areas.

Content of biliteracy: Textbooks on indigenous languages are mostly decontextualized and not considered authentic.

Media of biliteracy: Successive exposure to the Roman alphabet to access indigenous language literacy is more analogous to learning English as a foreign language than learning Chinese characters as a national language.

**2.1.1 CONTEXT OF BILITERACY.** The most desirable biliteracy in the Taiwanese society for the past two decades has been Chinese and English with the two writing systems, i.e., ideographic vs. alphabetic, clearly distinguished and equally valued. Other local languages are trying to gain status by standardization of scripts. However, the Chinese and Austronesian languages seem to have chosen different paths for orthography planning (Rau 1995a). Since a type of syllabic symbols (*bo-po-mo-fo*) has been used as a phonetic scaffolding for learning Chinese characters in Mandarin for over fifty years and has gained popularity as a major keyboarding system for computer literacy in Taiwan, any attempts to use Roman alphabets for orthography planning experienced tremendous difficulty until the last few years as English gained unprecedented popularity in Taiwan.

Other Chinese dialects (i.e., Southern Min and Hakka) have adopted a dual system with the Roman alphabet serving as phonetic symbols to facilitate language learning while Chinese characters are still used in formal writing. The Austronesian languages, on the other hand, have adopted only the Roman alphabet for orthography.

Compared to the socially dominant Chinese-English biliteracy communities in Taiwan, the Yami society on Orchid Island has only recently adopted the Roman alphabet for their orthography, since the publication of the New Testament in 1995. The older generation, above sixty years of age, tends to associate the Yami language with oral tradition, but is open to the promotion of literacy in Roman alphabets by Christian churches locally, whereas the younger generation is literate in Chinese characters only and is struggling to maintain their Yami language. As mentioned previously, with the exception of the village of Iraralay, the Yami teenagers no longer consider themselves proficient Yami speakers, according to the results of three sociolinguistic surveys over the past ten years by Rau

(1995b), Chen (1998), and Lin (2007). As a result, introduction of the Roman alphabet did not enhance Chinese-Yami biliteracy, although it has been viewed as conducive for English literacy.

While Internet infrastructure is well-established in Taiwanese society to support word processing and information search in Chinese and English, it is restricted to half of the villages on Orchid Island, where the language shift from Yami to Mandarin is most progressive.

In contrast to many other indigenous communities in the world as described in Hornberger (2003), the less “powerful”<sup>1</sup> communities in Taiwan are monolingual in a vernacular, whereas the more “powerful” communities are bi(multi)lingual, with one of them being the official language. As indicated in Rau (1995b), the most desirable language ability on Orchid Island is Chinese-Yami bilingual rather than either Chinese monolingual or Yami monolingual.

**2.2 DEVELOPMENT OF BILITERACY.** Children of the Taiwanese society are instructed to have productive skills in Chinese-English biliteracy. Chinese handwriting and word processing abilities are both necessary to function as an educated person in Taiwan. The development of English literacy (including English computer literacy) has followed the same development pattern of Chinese literacy by introducing the English alphabet early in pre-school and kindergarten. The model of teaching English to children in Taiwan actually parallels an English L1 language arts model. Native English speaking teachers are favored over non-native English teachers. Phonics is used to help children learn to read in English.

In Yami society, on the other hand, Chinese literacy is introduced as L1, whereas Yami literacy is taught in Chinese as an L2 subject for only one period of 40 minutes each week. As most children do not speak Yami as an L1, ethnic language teachers tend to emphasize receptive and oral skills. Yami orthography is downplayed, as authentic reading materials in Yami are lacking. In an Indigenous Language Proficiency Test indigenous junior high and high school students are required to pass in order to receive extra credit to enter the next level of schooling, the beginning level only tests listening and speaking abilities.

The development of biliteracy in Taiwan has the L1 and L2 models reversed for the less powerful and more powerful communities. In other words, instead of following an L2 model in teaching English as a second language to Chinese children in Taiwan, the addition of English literacy in the Taiwanese society tends to follow an L1 model. In addition, instead of teaching Yami or any other Austronesian languages as a first language in the indigenous communities, the addition of Yami literacy on Orchid Island tends to follow an L2 model.

**2.3. CONTENT OF BILITERACY.** The content of biliteracy in the Taiwanese society is geared toward the Chinese and increasingly American English majority literary culture, transmitted in multimedia on the Internet, whereas the Yami content of biliteracy focuses

---

<sup>1</sup> The term “powerful” follows Hornberger (2003), referring to politically, socially or linguistically dominant groups.



on the minority vernacular, which is highly influenced by Chinese and Japanese cultures. Children on Orchid Island primarily use the Internet for playing computer games rather than for building Yami literacy.

Once again, in contrast to Hornberger's model in which the content of biliteracy in the less powerful society is contextualized while it is decontextualized in the more powerful society, the Taiwanese case demonstrates the opposite. In other words, in the acquisition of English, the content is required to be "contextualized" and "authentic" for communicative needs. On the other hand, the content of online indigenous language textbooks for both elementary and junior high school students ([http://www.alcd.nccu.edu.tw/index\\_0.html%20](http://www.alcd.nccu.edu.tw/index_0.html%20)) is mostly "decontextualized," with direct translation from Chinese.

**2.4 MEDIA OF BILITERACY.** The media of biliteracy in the Taiwanese society have become increasingly similar and convergent in terms of their language structures and scripts, respectively, due to early exposure, or in many cases, simultaneous exposure to English. On the other hand, the successive exposure to Yami literacy is viewed as dissimilar and divergent from the Chinese structures and character writing to which they are exposed very early in life.

So far we have presented a critical evaluation of biliteracy continua of Chinese majority and Austronesian minority languages in Taiwan. This description has provided the context in which the e-Learning materials for the indigenous people in Taiwan are being developed.

One thing that is worth mentioning is that although the speakers of minority languages are not always positive towards the revitalization of their languages, nor are they always interested in participating in such activities, revitalization of the aboriginal cultural and linguistic inheritance of Taiwan is well supported by the Taiwanese governments, central and local.

**3. E-LEARNING PERCEPTIONS OF INDIGENOUS LANGUAGE TEACHERS IN TAIWAN.** In previous studies, the effectiveness of the e-Learning materials in Yami has been tested in a university classroom setting and assessed by Rau et al. (2007) and Yang (2007). However, no data were collected to reflect the attitudes of the indigenous people on the applicability of such a program to their individual community.

**3.1. CAN E-LEARNING CONTRIBUTE TO LANGUAGE REVITALIZATION?** In this section, we present data from a survey we conducted after a training workshop on literacy for indigenous language teachers in Taiwan to answer the question of local teachers' perceptions of the usefulness of e-Learning for language revitalization. Could the digital archiving of the endangered language be transformed into a useful tool for language revitalization? What are the indigenous teachers' impressions of the digital archive and e-Learning materials?

We conducted a pilot study consisting of informal interviews with several Yami participants at the Seminar on Language Revitalization on Orchid Island that we organized in April 2006. In that seminar, we asked a Yami tribal elder whether it is appropriate to use the web pages and animations to teach the Yami language. We received a very positive response from him. Therefore, the second author began to work with his students to produce

four episodes of Yami animation for our documentation project, the results of which will be discussed in Section 4. The opinions gathered from the Yami participants were further incorporated into the questionnaire on perceptions of e-Learning for indigenous languages that we designed for the current survey to explore if other indigenous groups have similar needs.

**3.2. INSTRUMENT AND PARTICIPANTS.** Our questionnaire (see Appendix 1) consists of ten question items, divided into four sections. The first section deals with the participants' attitudes toward the usefulness of computerized materials in language learning and the participants' knowledge of the Internet and e-Learning. The second section aims to determine their willingness to produce e-Learning materials with suitable assistance, and the type of materials they could produce. In the third section, the participants were asked if they would be willing to use e-Learning programs to teach their native languages. At the end of the third section, there was an open-ended question to discover potential problems of using e-Learning as teaching materials. The fourth section contains demographic information and self-evaluation of their native language ability.

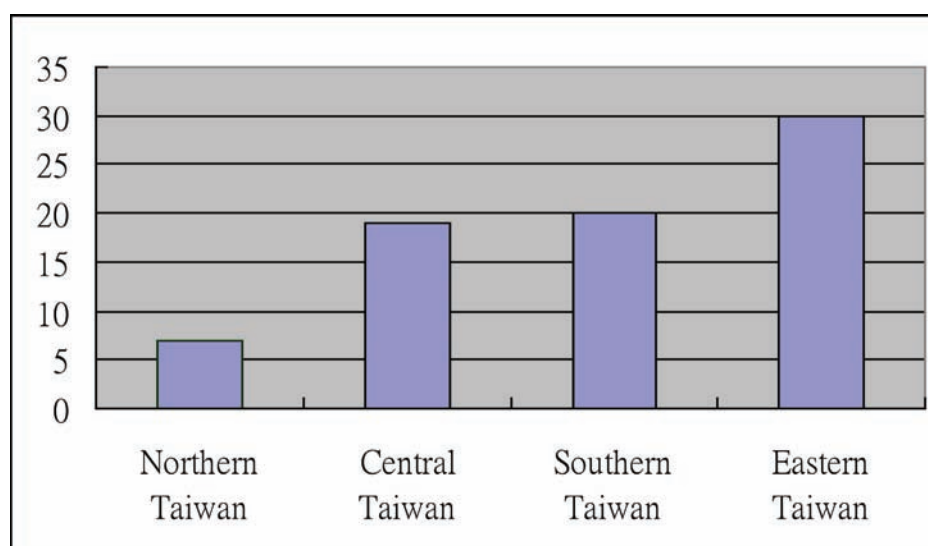


FIGURE 4. Geographic distribution of the participants

The questionnaire was administered to a group of 80 participating indigenous language teachers who were selected by the Council of Indigenous Peoples to represent all the indigenous language groups at a Training Workshop on literacy held at National Taiwan Normal University from November 22-24, 2006.

Before the survey, we gave an introduction to the Yami e-Learning program developed by our team at Providence University. In addition, a printed copy of our e-Learning program was distributed to each participant. We also asked permission and received consent from each to fill out the questionnaire.

All of the participants identified themselves as language teachers of their tribes. Seventy-nine participants turned in their questionnaires and 76 participants completed all questions. We used these 76 completed forms for data analysis.

The participants represented 14 language groups from 57 different villages. Five participants were living in cities away from home. The geographic distribution of their villages is shown in Figure 4, with the majority from Eastern Taiwan, followed by Southern and Central Taiwan. Less than 10% came from the Northern area. Over 60% of the participants responded that they were fluent in their native languages and had chances to speak daily.

### 3.3 SURVEY RESULTS.

**3.3.1 ATTITUDES TOWARD USEFULNESS OF COMPUTERIZED LEARNING MATERIALS IN LANGUAGE LEARNING.** Over 77% (N=59) of the teachers indicated that they agreed or strongly agreed that computerized learning materials are helpful in learning their native languages (Figure 5 and Table 1), while 17% (N=13) disagreed or strongly disagreed. Based on a cross-tabulation analysis, the teachers who expressed negative attitudes toward e-Learning also indicated that they had a problem with computer literacy.

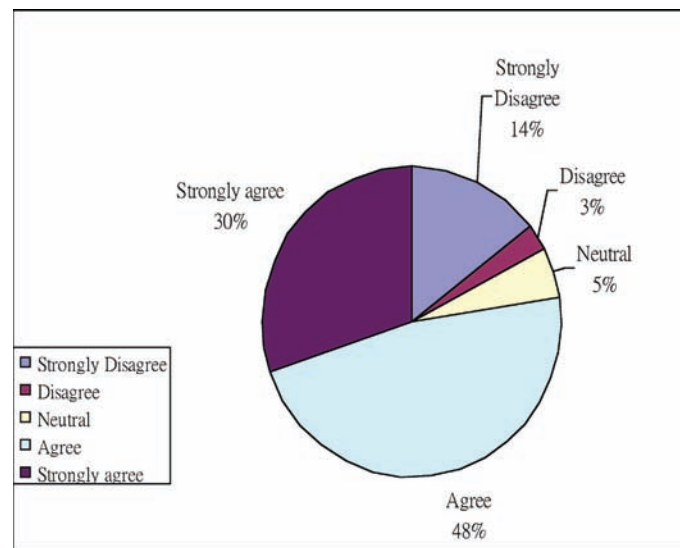


FIGURE 5. Distribution of categories from Table 1

TABLE 1. Computerized learning materials are helpful in learning the native language

Indicator	Number of Participants	Percentage
Strongly Agree	23	30%
Agree	36	47%
Neutral	4	5%
Disagree	2	3%
Strongly Disagree	11	15%
Total	76	100%

The majority of the participants (94%) also responded that computerized learning materials would be helpful for the young people to understand the traditional tribal cultures. Even members of the group which disagreed with the usefulness of e-Learning in language learning agreed that it is helpful for young people, as shown in Table 2.

TABLE 2 Computerized materials are helpful for the young people

Groups in Table 1	# of Participants/Total Participants in the group	Percentage
Strongly Agree	23/23	100%
Agree	34/36	94%
Neutral	2/4	50%
Disagree	2/2	100%
Strongly Disagree	11/11	100%

**3.3.2 TYPE OF MATERIALS THAT COULD ATTRACT YOUNG PEOPLE'S ATTENTION.** This section reports the participants' perception of the type of computerized learning materials that could attract their young people's attention. Among the five types of

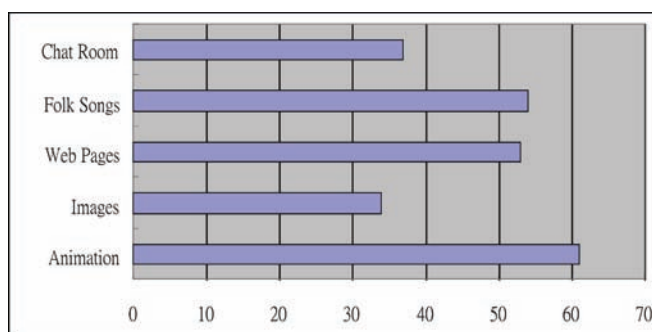


FIGURE 6: Type of learning materials that attract young people's attention

e-Learning materials listed (i.e., images, animation, web pages [text]<sup>2</sup>, folk songs and chat room), participants could choose any and all relevant types of learning materials which they considered could increase the interest in learning for their language. As shown in Figure 6, most participants picked animation, followed by folk songs and web pages. This result reveals that animation is an important topic in teaching young people to learn the indigenous languages.

**3.3.3 CONTENTS OF E-LEARNING.** When the indigenous language teachers were asked if they would be willing to try to create the contents of their own e-Learning programs with the help of the research team at Providence University, 90% (N=69) said yes, they would like to try to do this. Of the five categories of contents listed as choices to put in their future e-Learning programs (i.e., daily conversations, indigenous knowledge, ceremonies, legends, and basic skills), the top two selected by the most participants were

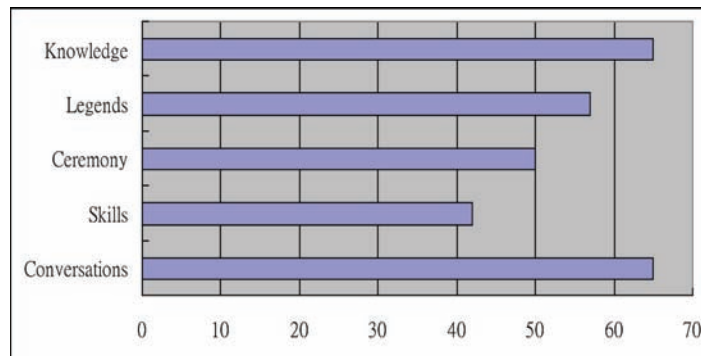


FIGURE 7: Favorite Contents of the e-Learning programs

indigenous knowledge and daily conversations, although the other three were also considered important by the majority of the participants. As illustrated in Figure 7, the majority of the tribal teachers consider that indigenous knowledge and daily conversations should be included in developing their own e-Learning programs.

**3.3.4 TEACHING INDIGENOUS LANGUAGES WITH E-LEARNING PROGRAMS.** When the participants were asked if they would be willing to use e-Learning programs to teach the indigenous languages should there be suitable e-Learning materials, 92% (N=70) responded “yes.” Furthermore, we asked them to indicate the types of content to be included in the e-learning programs that would be helpful for teaching. Among the five listed functions (i.e., grammar, online dictionary, demonstration of language use, online test, and online practices), the majority of teachers chose online dictionary and demonstration of language use, as shown in Figure 8.

<sup>2</sup> The original survey used the Chinese term for ‘web pages,’ which could include all the other categories, but our intent was to indicate text, which would have been clear to the respondents.

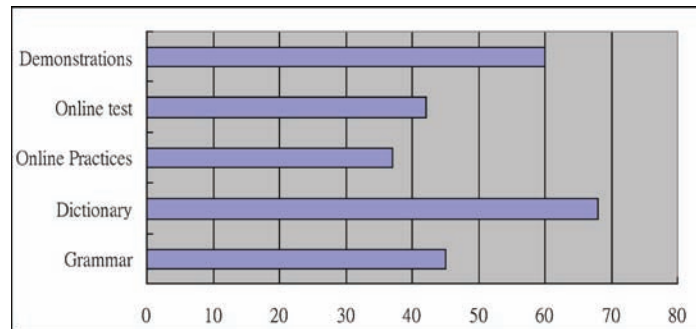


FIGURE 8: Type of e-Learning functions useful for language teaching

**3.3.5 TYPE OF SCHOOL MOST SUITABLE TO IMPLEMENT E-LEARNING.** Finally, we also asked what schools were most suitable to use the e-Learning materials. The majority responded that e-Learning could be implemented at all levels of schools in Taiwan, as shown in Figure 9.

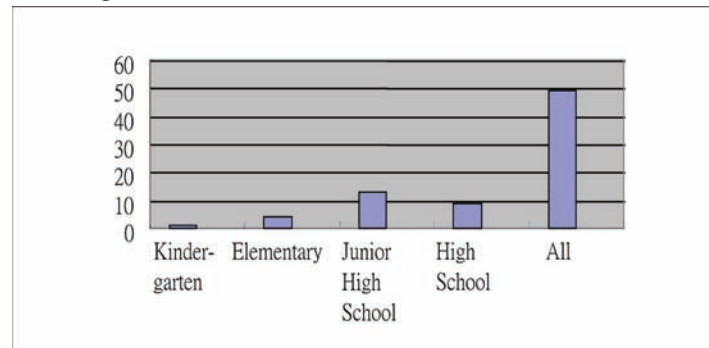


FIGURE 9: Which schools are suitable for use of the e-Learning materials?

**3.4 GEOGRAPHICAL DIFFERENCES.** To further analyze if geographical differences in Taiwan might contribute to different attitudes toward e-Learning, we used the following variables to build a correlation table:

- a. The value for geographic distribution of the indigenous teachers ranges from 1.1 to 4.4 based on living area, with weights for different townships. For example, if the teacher lives in northern Taiwan, s/he could have a weighted factor ranging from 1.1 to 1.4. This value was assigned the label 'a' for correlation analysis.
- b. The value assigned to the four statements "E-learning materials are helpful for learning indigenous languages," "E-learning materials are interesting to the young people," "I would like to try to create the e-learning materials with the assistance

from the research team at Providence University,” and “I would like to try to use e-learning materials for teaching” ranged from 0 to 1 depending on their answer; a positive answer being 1.0 and a negative answer 0, with the weight adjusted according to their self-description of language fluency. These values were assigned the labels b-e, respectively, for correlation analysis.

The correlation table (Pearson  $r$ ) is shown in Table 3. We found that the  $r$  of a\*d is much weaker than the others. This indicates that indigenous teachers in different areas have different attitudes toward creating their own e-learning materials. This can be attributed to the sign of ‘digital divide’ in eastern Taiwan, where access to the Internet is far more limited than in other sections of Taiwan.

TABLE 3 Correlations based on the geographic distribution

Variables	Pearson correlation coefficient	df
a*b	0.403	76
a*c	0.560	74
a*d	0.082	76
a*e	0.608	75

**3.5 DISCUSSION.** The results of the survey indicate that developing e-Learning materials for indigenous language teaching and learning is highly acceptable and strongly supported by the indigenous language teachers in Taiwan. It can be implemented at all levels of schools in Taiwan. However, the problem of the digital divide in eastern Taiwan needs to be addressed before e-Learning programs can be launched in all indigenous communities.

The majority of teachers still perceived e-Learning as most useful for providing resources, e.g., providing texts, dictionary, and grammar rather than using it for online practice and online tests. This attitude indicates that training workshops on how to use e-Learning materials should be conducted before the advantages of e-Learning can be fully realized. Furthermore, it is important to publish indigenous language dictionaries online and transform traditional textbooks containing conversations, grammatical explanations, and cultural information into multimedia format online. This indicates the importance of building a team consisting of a linguist, an IT professional and an indigenous language teacher to establish such a program, as proposed by Yang and Rau (2005) and Rau and Yang (forthcoming 2008).

Finally, animation based on indigenous languages and cultures should be widely exploited for language revitalization as it is considered the best tool to attract the attention of young indigenous people to learn their own languages. This will be discussed further in Section 4.

#### 4. ATTITUDES OF THE YAMI COMMUNITY ON ORCHID ISLAND TOWARD E-LEARNING.

**4.1 FEEDBACK FROM YAMI TEENAGERS.** In this section, we summarize results from two recent MA theses supervised by the authors. The results of a qualitative study

on the Yami teenagers' attitudes toward the animation productions will be discussed in the next section. First, a word on Yami teenagers' use of Internet is in order.

When we visited Orchid Island in April 2006, we found that Yami students on the island are just as attracted to the Internet as are Chinese students in Taiwan. However, most of the web pages they visited are in Chinese. There are very few Yami websites. This observation inspired us to investigate the teenagers' attitudes toward the Yami language websites we developed, when they finally had a chance to view them. This specific question was investigated in the two MA theses by Lin (2007) and Yang (2007) in their studies on Yami language vitality and Yami e-Learning, respectively.

According to Lin (2007), setting up a Yami documentation website was ranked by 117 Yami teenagers as the third most efficient way for maintaining Yami language, preceded by encouraging Yami people to speak the language and promotion of ethnic language education at school. They expressed equally high positive attitudes toward our Yami documentation and e-Learning websites regardless of their Yami language proficiency. In general, they were more interested in the Yami e-Learning websites than the documentation one because they found the former useful for learning Yami following the easy-to-difficult materials.

Yang (2007) in her needs analysis of our e-Learning program found that the third year junior high school students on Orchid Island had more positive evaluation of our websites than did younger students. She claimed that the reason might be that they had pressure to pass the test of "certification of indigenous languages" so that they could get the 35% bonus points in the Basic Competence Test, a High School Entrance Examination. Beyond the current online indigenous language textbooks for both elementary and junior high school students, endorsed by the Ministry of Education,<sup>3</sup> our websites are probably the only resources available for test preparation covering all four language skills in Yami (listening, speaking, reading and writing). Furthermore, Iraralay teenagers, the only competent young Yami speakers on the island, expressed more willingness to use the e-Learning materials to study after school than those in other villages. Overall, the teenagers suggested that animation, films, and pictures would be the three elements on a website most likely to attract their attention. They expressed that the greatest learning difficulties they have are Yami vocabulary and spelling and suggested that our website add two sections, (1) introducing vocabulary with pictures and (2) providing dictation and typing games to practice spelling.

Since animation was mentioned by both indigenous language teachers and Yami teenagers as the most desirable component in e-Learning, we instructed a group of undergraduate computer-science majors to develop a set of animations to accompany each lesson and exercise of our e-Learning program. They were created under the instruction of the authors. The students were first given information about the Yami language and culture before they were asked to design six cartoon-style characters for the animation. All the characters speak Yami. The animations were made using Flash MX and ranged in duration from 30

---

<sup>3</sup> The online indigenous language textbooks for both elementary and junior high school students are available from ([http://www.alcd.nccu.edu.tw/index\\_0.html%20](http://www.alcd.nccu.edu.tw/index_0.html%20)), the contents of which are mostly "decontextualized" with direct translation from Chinese



seconds to 1 minute and 30 seconds. These animation productions were taken to the high school on Orchid Island by two project team members in January 2007. Six Yami students were interviewed and asked to express their opinions.

The two authors also instructed a second group of computer science majors to create four animated stories based on the famous Yami legends of Mr. Paloy and organized two animation dubbing contests, in Yami and English. The following sections describe the feedback of these initiatives.

#### **4.2 FEEDBACK ON ANIMATION AND ITS USE IN LANGUAGE LEARNING.**

The interviews with six high school Yami students were conducted in two stages. First, they were instructed to watch at least three of the four Yami animation productions (<http://yamiproject.cs.pu.edu.tw/yami/corpus7.htm>). After they finished, each student was asked to fill out a questionnaire with six questions (see Appendix 2) related to their impressions of these animated characters and the usefulness of animation in Yami language learning.

The results indicate that most students agreed that animation would be helpful for them to learn the Yami language. Four students responded that they fully accepted the style of presentation in the animation while the other two responded that although the animation did not match the local style, they could still accept it. However, when the students were asked if Yami living and cultural items could be transformed to animation to reflect authentic Yami lifestyle, their opinions were divided. Three said yes but the other three had reservations, saying that they needed to see the film production first before they could judge.

Overall, these students expressed positive attitudes towards the animation. We asked them to choose one character that they liked and one that they disliked in the animation. These students picked out two characters that they disliked. One character had darker skin than the Yami people and the other had a timid voice. In addition, two students said they hoped the characters in future animations could speak more naturally, like real conversations.

#### **4.3 FEEDBACK ON ANIMATION DUBBING CONTEST ON THE LEGENDS OF**

**MR. PALOY.** The second set of animations was used as a tool for language revitalization and raising visibility of an endangered language. In response to the Yami elders' call for developing animated stories about their culture, we organized an animation dubbing contest on April 28, 2007 on Orchid Island. Although our intention was to attract teenagers to form teams to compete in this event, we actually succeeded in scaring everyone away from this because the adults were afraid of advanced technology and the youngsters were afraid of speaking Yami. We finally decided to invite the four judges, a friend of a judge and our Yami project consultant Maa-neu Dong to form two teams, with three people in each group dubbing two stories in Yami. One judge at the beginning was reluctant to try it and insisted that the legends of Mr. Paloy were only told by the Iratay and Imowrod villages, so she could not even make up the story lines. However, after she joined a team and tried a few times, she completely changed her mind.

They first watched the animated stories several times. One chose to read the Chinese translations of the Yami scripts that Dong had written for the four stories. After practicing a few times with their own team, they gradually realized it was only a simple role-play of a locally familiar theme. Finally, each team finished recording two trials of each of the two

animated legends within half an hour. When they finished the task, they sat around to reflect on their impressions. They were surprised by their positive attitudes toward this creative effort. One of the judges, who had been in charge of implementing language revitalization initiatives in Iraralay, decided to incorporate this tool into future Yami speech contests.

Meanwhile, we also organized an English animation dubbing contest in the Department of English Language, Literature, and Linguistics at Providence University as a special part of the annual English festival, this year featuring endangered language revitalization. Using the same animation of Mr. Paloy and applying the same regulations, we attracted five teams to sign up, mostly freshmen and sophomores. The participants prepared the English script for their assigned animation ahead of time, either individually or as a group. The contest was held on campus on May 17, 2007. The participants role-played in English to the animation with Chinese subtitles by reading their own scripts and playing their sound effects in front of the audience and three judges. Two awards were given based on their performance in English. This event has stimulated great interest within the department and has become integrated into the program of future English festivals.

**4.4 DISCUSSION.** The animation dubbing contest was found to be a potentially useful tool for language revitalization because it fits the models of biliteracy in both communities. Dubbing in Yami matches the monolingual oral tradition of the Yami community and thus will attract attention from both adults and children, while reading the English script to dub the animation with Chinese subtitles matches the biliteracy in the Taiwanese community and also increases visibility of the language and culture of an endangered language.

An effective way to develop useful animation for indigenous language revitalization is for indigenous language teachers to collaborate with computer science teachers. After a set of animations are created to fit the script of a popular indigenous legend, a dubbing contest can be held in the community to select the best version or versions to go online.

We expect a polished product of animation in an indigenous language with subtitles in either Chinese or English will attract not only indigenous children who spend long hours playing computer games in their villages but also Taiwanese children who are eager to learn English by watching cartoons. The increasing visibility of an indigenous language on the Internet will lead indirectly to language revitalization.

**5. CONCLUSION AND SUGGESTIONS FOR DEVELOPMENT OF E-LEARNING PROGRAMS.** This study is a summative assessment of the usefulness of e-Learning in language documentation and revitalization. We began with an analysis of the power relations according to the continua model in Taiwan and Yami communities. We found Yami literacy was analogous to foreign language acquisition for Yami teenagers. The only impetus for them to use our e-Learning materials currently is for test preparation.

Although the attitudes among the indigenous language teachers toward creating e-Learning materials with the help of an experienced team were generally positive, a digital divide is still apparent on the east coast of Taiwan. Overall, most indigenous language teachers were not familiar with new technology and required technical assistance.

We also found that the existing e-Learning tools can be useful for technology-integrated teaching programs. Of all the available tools, the indigenous language teachers showed most interest in using the (1) online dictionary, (2) online audio or video clips on daily conversations and indigenous knowledge, and (3) online animation in developing their own course materials.

The online dictionary should have a Chinese interface, a dynamic visual presentation with rich sounds and images of phonetic and text explanations for each vocabulary item, and intelligent search functions that can assist teachers and students to find the correct words or phrases. Furthermore, in addition to being a collection of the most frequently used words, suitable for a general audience, the online dictionary can also be customized to fit the needs of individual teachers, as described in Yang et al. (2007).

The online audio and video clips in the archives can be edited for language documentation. A multimedia database system makes it possible to search for suitable topics in the clips, thus allowing language teachers to create their own teaching materials and conduct online teaching.

Finally, the online animations can provide a virtual classroom, and are an effective way of presenting stories with sound tracks. It is necessary to develop a user-friendly computer platform to allow indigenous language teachers to transform their own texts into animation. Animation was identified by both indigenous language teachers and Yami teenagers as the most desirable element in e-Learning. Attempts to use animation dubbing contests as a tool for language revitalization were found to be promising, as they match the models of continua of biliteracy.

**APPENDIX 1**

Questionnaire on e-Learning and Creating Digital Materials for Teaching Austronesian Languages in Taiwan

1. E-learning materials are helpful for learning indigenous languages.  
 Strongly Agree.  Agree.  Neutral.  Disagree.  Strongly Disagree.
  2. Which digital format is most suitable to attract the attention of the younger generation of your people?  
 CD Titles.  Web Pages.
  3. E-learning materials are interesting to the younger generation of your people.  
 Yes, they are helpful.  No, they are irrelevant.
  4. What type of e-learning materials can attract your young people's attention? (Choose all that apply)  
 Images.  Animation.  Web pages.  Folk Songs.  Chat Room.
  5. I would like to try to create e-learning materials with assistance from the research team at Providence University.  
 Yes, I am willing to try.  No, it is probably too difficult for me.
  6. What contents should be put in the indigenous language e-learning programs? (Choose all that apply)  
 Daily conversations.  Basic skills.  Ceremonies.  Legends.  Indigenous knowledge.
  7. In what schools is it most suitable to use e-Learning materials? (Choose all that apply)  
 Kindergarten.  Elementary school.  Junior high school.  Senior high school.
  8. Which functions should be added in the e-learning programs for indigenous language teaching and learning? (Choose all that apply)  
 Grammar.  Dictionary.  Demonstration of language use.  Online tests.  
 Online practice.
  9. I would be willing to use e-Learning programs to teach the indigenous language, should there be suitable e-Learning materials.  
 Yes.  No.
  10. What are potential problems of using e-Learning materials?
-

Background Information

1. What is your ethnicity? \_\_\_\_\_ In what county, township, and village do you live?  
\_\_\_\_\_ Do you live in an urban area? \_\_\_\_\_
2. Are you an indigenous language teacher? \_\_\_\_ Yes \_\_\_\_ No
3. Do you use your indigenous language everyday? Please explain your language use.  
\_\_\_\_\_

**APPENDIX 2**

Questionnaire on the use of animation for Yami language learning

Please watch at least three animation productions and answer the following questions.

1. Could animation help you learn the Yami language?

No.  Yes, It is helpful.  Yes, It can increase my motivation to learn Yami.

2. Is the style of presentation in the animation appropriate for the Yami language learning materials?

It is appropriate.  It is not appropriate.

3. Do you like the characters in the animation productions?

Yes, I do.  No, I do not.

4. Who is your favorite character?

I like \_\_\_\_\_. The reason is \_\_\_\_\_.

5. Which characters do you dislike?

I don't like \_\_\_\_\_. The reason is \_\_\_\_\_.

6. How can these animation productions be improved? Can Yami living and cultural items be transformed to animation to reflect authentic Yami lifestyle?

## REFERENCES

- CHEN, HUI-PING. 1998. A sociolinguistic study of second language proficiency, language use, and language attitude among the Yami in Lanyu. MA thesis, Providence University, Taiwan.
- CSATÓ, EVA A. and DAVID NATHAN. 2003. Multimedia and documentation of endangered languages. In *Language Documentation and Description*, ed. by Peter K. Austin, 1:73-84. London: SOAS.
- DONG, MAA-NEU and D. VICTORIA RAU. 2000. Yami textbook (in Chinese). Taipei: Crane.
- EISENLOHR, PATRICK. 2004. Language revitalization and new technologies: Cultures of electronic mediation and the refiguring of communities, *Annual Review Anthropology* 33:21-45.
- FITZGERALD, MICHAEL and ROBERT DEBSKI. 2006 Internet use of Polish by Polish Melburnians: Implications for maintenance and teaching, *Language Learning and Technology*, 10.1:87-109.
- HORNBERGER, NANCY H. 2003. Continua of biliteracy: An ecological framework for educational policy, research, and practice in multilingual settings. Clevedon: *Multilingual Matters*.
- HORNBERGER, NANCY H. and ELLEN SKILTON-SYLVESTER. 2003. Revisiting the continua of biliteracy: International and critical perspectives. In *Continua of biliteracy: An ecological framework for educational policy, research, and practice in multilingual settings*, ed. by Nancy H. Hornberger, 35-67. Clevedon: *Multilingual Matters*.
- LIN, YI-HUI. 2007. A sociolinguistic study on Yami language vitality and maintenance. MA thesis, Providence University, Taiwan.
- LIOU, HSIEN-CHIN. 2005. Computer-assisted language learning in East Asia. In *Encyclopedia of Distance Learning*, eds. by Caroline Howard, Judith V. Boettcher, Lorraine Justice, Karen Schenk, Patricia Rogers, and Garry A. Berg, 341-351. Hershey, PA: Idea Group Inc.
- PARKS ASSOCIATION. 2005. Taiwan and Korea lead in the adoption of consumer technologies. Retrieved on May 25, 2007: [http://www.parksassociates.com/press/press\\_releases/2005/gdl6.html](http://www.parksassociates.com/press/press_releases/2005/gdl6.html)
- RAU, D. VICTORIA. 1995a. The scientific and social principles for orthography development: a case study of Atayal (in Chinese), *Proceedings of Conference on Taiwan Austronesian Language Maintenance and Development*, Ministry of Education of R.O.C., 31-47.
- RAU, D. VICTORIA. 1995b. Yami language vitality (in Chinese), Paper presented at the Conference on Language Use and Ethnic Identity, Institute of Ethnology, 6/16, in Academia Sinica, Taipei
- RAU, D. VICTORIA, HUI-HUAN CHANG, and MAA-NEU DONG (to appear). A tale of two diphthongs in an indigenous minority language. In *Variationist Approaches to Indigenous Minority Languages*, Eds. by James N. Stanford and Dennis R. Preston. Amsterdam: John Benjamins.
- RAU, D. VICTORIA and MENG-CHIEN YANG. (to appear). Digital transmission of language and culture. In *Endangered Languages of Austronesia*, ed. by Margaret Florey. Oxford University Press.
- RAU, D. VICTORIA, HUI-HUAN CHANG, YING-SHENG TAI, ZHEN-YI YANG, YI-HUI LIN, CHIA-CHI YANG, and MAA-NEU DONG. 2007. Teaching and Learning an endangered Austronesian language in Taiwan. Paper presented at the International Conference on Austronesian Endangered Language Documentation, June 5-6, in Providence University, Taichung, Taiwan.

- RAU, D. VICTORIA, MENG-CHIEN YANG, and MAA-NEU DONG. 2007. Endangered language documentation and transmission. *Journal of National Council of Less Commonly Taught Languages (NCOLCTL)*. University of Wisconsin at Madison, 53-76.
- RAU, D. VICTORIA and MAA-NEU DONG. 2006. *Yami Texts with Reference Grammar and Vocabulary*. Language and Linguistics, Institute of Linguistics, Academia Sinica, Taipei, Monograph A-10.
- RAU, D. VICTORIA, MAA-NEU DONG, MENG-YING LIN, HUI-HUAN CHANG, and YONG-CHANG HSU. 2005. *Multimedia Materials of Yami Language*, Technical Report, Department of English Language, Literature and Linguistics, Providence University.
- VILLA, DANIEL J. 2002. Integrating technology into minority language preservation and teaching efforts: An inside job, *Language Learning and Technology*, 6.2:92-101.
- WARD, MONICA and JOSEF VAN GENABITH. 2003. CALL for endangered languages: Challenges and rewards. *Computer Assisted Language Learning* 16.2-3:233-258.
- WEF (WORLD ECONOMIC FORUM). 2007. *Global information technology report 2006-2007*. Retrieved on May 25, 2007: <http://www.weforum.org/en/initiatives/gcp/Global%20Information%20Technology%20Report/index.htm>
- YANG, CHIA-CHI. 2007. *Understanding users' perceptions: Using needs analysis to developing Yami language website*. MA thesis, Providence University, Taiwan.
- YANG, MENG-CHIEN and D. VICTORIA RAU. 2005. *An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan*. Paper presented at the ALR-2005, Jeju Island, Korea.
- YANG, MENG-CHIEN, HSIN-TA CHOU, HUI-SHIUAN GUO, and GIA-PYNG CHEN. 2007. *On Designing the Formosan Multimedia Word Dictionaries by a Participatory Process*. Paper presented at the International Conference on Austronesian Endangered Language Documentation, June 5-6, in Providence University, Taichung, Taiwan.

D. Victoria Rau  
dhrau@pu.edu.tw

Meng-Chien Yang  
mcyang2@pu.edu.tw



# 8

## Indigenous Language-Informed Participatory Policy in Taiwan: A Socio-Political Perspective

Yih-Ren Lin<sup>°</sup>, Lahuy Icyeh<sup>°</sup> and Da-Wei Kuan (Daya)<sup>†</sup>

<sup>°</sup>*Providence University*

<sup>†</sup>*University of Hawai'i at Mānoa*

This chapter highlights the importance of incorporating indigenous language and its daily practice in the local context of newly transformed indigenous policy in Taiwan. Currently, the official indigenous people's language policy is relatively confined to curriculum development and certification of indigenous peoples' language abilities with little consideration of language practices in real socio-political situations. This paper questions whether the revitalization of endangered indigenous languages can rely only on language policy *per se*. The participatory action research (PAR) methodology is employed as a main research method in inhabited Atayal communities. This chapter is divided into three main parts: firstly, a brief socio-political history of indigenous people in Taiwan is provided; secondly, two socio-political official projects related to traditional territory sovereignty are analyzed: their failure is revealed due to the neglect of indigenous language and local participation; thirdly, a case from an Atayal village, *Smangus*, is provided to show how indigenous languages can be revitalized through combining the villagers' daily practices and participation. In conclusion, this chapter argues for a combining of language policy with other socio-political policies so as to create environments in which indigenous peoples can speak their own languages.

**1. INTRODUCTION.** This chapter<sup>1</sup> highlights the importance of incorporating indigenous languages and daily practices into Taiwan's policies for indigenous peoples. Since 2000, the authors have been involved in several projects under the "New Partnership Policy." The policy originated from a commitment by Mr. Shui-Bian Chen to representatives of indigenous tribes when he was running for his presidency in 1999. This policy has become President Chen's principal governmental policy for indigenous peoples when he was elected in 2000. The commitment has the following seven goals: 1) recognize the natural rights of indigenous peoples in Taiwan; 2) promote the autonomy of indigenous peoples; 3) sign land treaties with indigenous peoples; 4) recover the traditional names of indigenous communities and living places; 5) recognize the traditional territories of indigenous peoples; 6) allow traditional uses of natural resources and indigenous peoples' autonomic development; and, 7) achieve equal representation of indigenous peoples in parliament. The policies of the central government's Council of Indigenous Peoples have been directed to work toward achieving these goals in line with the president's commitment.

---

<sup>1</sup> The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC\_94-2621-Z-126-001. Special thanks will also go to the Atayal villagers with whom we have been working and from whom we have been learning, especially the *Smangus* villagers.

The traditional territory mapping project, documentation of traditional ecological knowledge and co-management of natural resource management—with which the authors are involved—have been regarded as flagship projects. This involvement indicates that the importance of indigenous language practices is rarely noticed in government socio-political projects even though indigenous languages typically play a crucial role in revitalizing traditional cultures.<sup>2</sup> Rather, these projects, carried out primarily by the Council of Indigenous Peoples, utilized a top-down technocratic operational framework that offered little opportunity for local participation. This chapter argues that the major reason for endangerment of Taiwan's indigenous languages is the alienation of indigenous languages from their socio-political contexts, which may have a greater influence on contemporary indigenous communities than language policy. In other words, we argue that revitalization of endangered indigenous languages not only relies on language policy, but also requires combining language policies with other socio-political policies to create environments in which indigenous peoples can speak their own languages.

This chapter brings together both indigenous and non-indigenous voices. The first author, a non-indigenous ecological scholar<sup>3</sup>, has been involved substantially in major events such as proposed Maqaw national park and traditional territory mapping.<sup>4</sup> The second and third authors are Atayal and are committed to revitalization of their culture. Participatory action research (PAR) is employed as the principal methodology in our research (Whyte 1991; Stringer 1996). In the context of indigenous peoples' studies, this methodology has only recently been applied in Taiwan (Lin 2005). Different from conventional anthropological and linguistic studies, this methodological scheme does not entertain a clear distinction between researchers and those researched, but rather puts relatively more importance on reflection and capacity-building of those researched from a researcher's perspective. Therefore, one evaluation criterion is whether a researcher can explain explicitly the entire research process and strategies to generate trustworthy data. Stringer (1996) indicates that understanding is generated from a dialectical circular process of researcher's observation, thinking, and actions via their participation and interaction with research subjects. How-

---

<sup>2</sup> Although the New Partnership Policy mainly focuses on curriculum development and certification of indigenous peoples' language abilities, indigenous language speaking ability in its daily life practices should be considered crucial to the policy's success (Sun 2000, 2005). The bureaucracy has not created relationships with other socio-political policies in terms of overall cultural revitalization.

<sup>3</sup> The first author is a committee member of the "Maqaw National Park Advisory Committee" and used to be a key member of the Project of Indigenous Traditional Territory Survey. Both activities are described later in this chapter.

<sup>4</sup> Maqaw (*Litsea cubeba*), an Atayal word, is a fragrant plant used by Atayal people for seasoning foods. However, the term "Maqaw" in the context of establishing a new national park has multiple definitions: (i) a seasoning plant; (ii) the Chi-Lan Shan area where the most precious and contested Chaemacypris forest is located—one mountain top in this area is rich with this tree; and (iii) the name of the proposed and controversial national park. This issue has been markedly promoted by the media through the social movement of saving ancient Cypress forest (Lee 2004).

ever, regardless of the type of action research (Castellanet and Jordan 2002)<sup>5</sup>, the question “Who is participating?” must be asked. Additionally, one must identify whether the researcher or local actors determine the direction of the research project.

Based on methodological concerns, this chapter describes the critical processes in projects related to transformative policies for indigenous peoples with which the authors have been involved since 2000, and discusses the implications of a lack of indigenous language input. Moreover, the two Atayal authors, who are fluent in the Atayal language and have considerable commitment to Atayal culture, play an important role in evaluating and analyzing the data from an insider perspective. From 2000 to 2004, discursive data was collected through participant observations, documentation, workshops, meetings, cultural teaching activities and individual interviews with elders from major events of proposed Maqaw national park and traditional territory mapping. Some of our analytical results have been published in Chinese (mainly) and English (Lin 2004, 2005, Lin *et al.* 2006) from an ecological perspective; however, the language issue was not addressed. The issue of language is discussed in detail in this chapter.

The remainder of this chapter is organized as follows. First, a brief history of the socio-political situation of Taiwan’s indigenous people is provided and contrasted with the ideals in the “New Partnership Policy” from an Atayal perspective. Second, two cases are introduced that are related to the Policy’s so-called “partnership” and construction concerns as they are linked to local ecological knowledge, namely, i) the proposal to establish a national park based on co-management by indigenous communities and the state, and ii) the project of mapping the traditional territories of indigenous peoples. Moreover, the lack of indigenous language input is discussed. Third, a participatory case in which the authors were involved suggests a way of creating a socio-cultural field for revitalizing the languages of indigenous peoples. This chapter concludes with a recommendation that linguistic input into Taiwan’s current socio-political policies for indigenous peoples is demanded.

**2. A BRIEF SOCIO-POLITICAL HISTORY OF INDIGENOUS PEOPLE IN TAIWAN.** Taiwan is an island of Austronesian indigenous peoples. From the 17<sup>th</sup> to 19<sup>th</sup> centuries, the plains in western Taiwan were governed by the Dutch, an exiled government of the Ming Dynasty and then by the Ching Dynasty. However, the plains in eastern Taiwan and central mountains, which are home to indigenous tribes, were not governed by foreign governments until 1895, which was the beginning of the Japanese colonial era. The Japanese colonial government implemented a land survey in 1898, and then in 1910 initiated a five-year military project to conquer indigenous peoples in Taiwan. The mountainous areas previously “owned” by different indigenous communities were then nationalized. In 1925, the National Forestry Survey Project confined indigenous people to Reserved Lands, which were small and fragmentary land parcels in the mountains. At the same time, many communities were forced to migrate to low mountainous areas, and shift from traditional hunting and gathering to agricultural production.

---

<sup>5</sup> In their research of the PAR methodology, Castellanet and Jordan (2002) identified three research orientations based on the extent of researcher participation in the research process.

In 1945, the Kuomintang<sup>6</sup> (KMT) government replaced the Japanese colonial government in Taiwan. The KMT government retained the Reserved Lands Policy and exploited forestry resources. Taiwan's indigenous peoples resisted this domination just as they had previously dealt with colonists in different historical periods. Many bloody battles between the Japanese military and indigenous peoples occurred. After WW II, the indigenous intelligentsia asserted the right to self-determination. However, many were arrested and executed by the KMT government.

During the late 1980s and 1990s, while Taiwan was going through democratic reforms, many demonstrations and protests were held over land rights, autonomy and self-interpretation of indigenous culture and history. One significant achievement during this period was that the congress incorporated the term "indigenous peoples" into the Constitution in 1994. Before that, indigenous people were called "mountain people" in the Constitution. In 1997, the central government created the Council for Indigenous Affairs. However, indigenous land rights were rarely discussed.

The KMT's policy toward indigenous peoples was largely devoid of recognition of indigenous rights; thus, discussing what "traditional territory" means is very difficult. However, in the 2000 presidential campaign, presidential candidate Mr. Chen Shui-Bian (president at the time of writing) announced a "New Partnership Policy" as his major indigenous policy. This policy, which committed the government to recognizing indigenous claims to traditional territories, was codified in legislation when Chen assumed the presidency.

The following presents two cases that are extremely relevant to President Chen's New Partnership Policy. Thus, this chapter also discusses the limited understanding of the role of indigenous languages and peoples in these two cases.

### 3. TWO CASES RELATED TO THE "NEW PARTNERSHIP POLICY."

**3.1 CASE I: LACK OF NATIVE LANGUAGE INPUT — CO-MANAGEMENT OF THE PROPOSED MAQAW NATIONAL PARK.** The controversy surrounding establishment of a co-managed national park is worthy of discussion and was initiated by a social movement of nature conservationists to save ancient Cypress forest (Nokan 2003; Lee 2004; Lin 2004). This controversial issue was the first government project to involve indigenous peoples in natural resource management, and has generated considerable academic discussion in Taiwan. Over 20 doctoral and master's theses are related to this issue. Additionally, the Maqaw national park issue is utilized in a popular international conservation textbook as a case study illustrating multiple environmental values in a contesting context (Lin 2005). Development of this issue is strongly related to traditional ecological knowledge and community development of Taiwan's indigenous peoples. Cypress trees (*Chaemacyparis* spp.) in Taiwan belong to the same *Cupressaceae* family as *Cupressus* spp. in the southeastern swamplands in the United States. These trees in Taiwan are found at elevations of 1800–2500m, can grow as tall as 60m tall and to 20m in girth. Based on habitat, ecology, commercial value, and appeal, these trees resemble California redwoods. However, after nearly a century of intensive logging sanctioned by several governments, only two large Cypress forests remain in Taiwan—Chi-Lan Shan dominated by *C. obtusa* var. *formosana* and Sho-Ku-Luan Shan dominated by *C. formosensis*. With de-

<sup>6</sup> The nationalist political party.

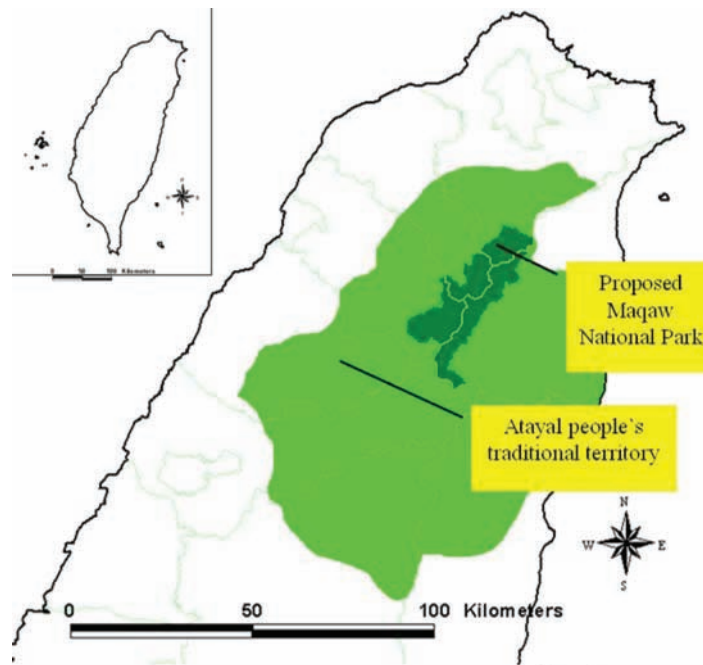


FIGURE 1: The Map of the proposed Maqaw National Park within Atayal People's Traditional Territory

mocratization of the political landscape and increased ecological literacy, a “Ban for Logging Natural Forests” was instituted in 1991 by the Council of Agriculture—a cabinet-level government department—in response to environmental concerns (Huang 2004). However, many in the forestry industry still pursued access to cypress timber through political means due to the economic value of cypress trees. Since 1986, the Council of Veterans (COV), another cabinet-level department, has “salvaged” cypress trees blown over by high winds and killed by lightning strikes and disease, and replanted young cypress trees in openings. In 1998, the COV again proposed a new 5-year plan to remove dead trees in most of the remaining cypress forests, with the rationale that if they do not, the living forests will no longer be able to reproduce themselves and will eventually die out.

As a result, the COV’s proposal provoked a number of nature conservationists including academics, legislators, and some local environmental groups<sup>7</sup> that subsequently banded together and launched “The Rescue Cypress Forests Movement League”. The group’s goal was to force the COV to halt its salvage operations and revoke its right to manage the cypress forests. On Dec. 25, 1999, the league held a rally with the goal of establishing a new national park for protecting ancient Cypress forest. This rally was staged during Taiwan’s presidential campaign. To obtain the support of environmentalists, the Democratic Pro-

<sup>7</sup> Notably, no indigenous peoples were involved in the beginning of this event.

gressive Party (DPP) candidate Mr. Shui-Bian Chen responded by promising to establish a new national park if elected president. This new national park, however, would have to be carved out of the traditional Atayal territory. The Atayal comprise the second largest indigenous group in Taiwan. Figure 1 shows the area of the new national park within the traditional Atayal territory.

The new national park proposed by the league therefore introduced a new interest group and set of values into the cypress forest conservation debate, namely, the Atayal. Taiwan already has six national parks comprising 8% of the island's total area. When these parks were created, indigenous peoples were frequently ignored and dispossessed (Sung 1999). Understandably, the Atayal tended to oppose any additions to the national park system. Conversely, indigenous peoples are not natural allies of the COV—which also opposes creating new national parks—because traditional subsistence uses in the highlands do not include logging, nor is logging consistent with those uses. After discussion with Atayal opinion leaders, the league decided to include local indigenous people in the design process for the new national park, and to establish a new “co-management” protocol for the new national park that would have indigenous people on the park's board of directors and employ indigenous people as, for example, park rangers, guides, and naturalists. Thus, the proposed new national park would preserve the precious old-growth cypress forest, revitalize indigenous culture and language, and improve the livelihood of indigenous people. The national park proposal would also need to know the Atayal's traditional ecological knowledge of protecting the forest in their own way.

In 2000, Mr. Shui-Bian Chen won the presidency. The league began lobbying the new president to make good on his promise to create a new national park. In October 2000, the league and some indigenous groups appealed to the Ministry of the Interior—which oversees national parks—to respect their joint concerns. In response, the Maqaw National Park Advisory Committee was formed. Unavoidably, the proposed new national park was also valued highly by the central government and put on its political agenda as one means of realizing Chen's “New Partnership Policy.” Over the course of more than 10 meetings, the committee discussed issues related to co-management, means of implementation, and national park boundaries. The committee's main purpose was to establish the new national park, preserve the cypress forests, and benefit local indigenous people. This committee soon became a forum for dialogue between conservationists, indigenous peoples, scholars, and the state. However, the committee has not succeeded in eliminating skepticism and distrust among opposing groups. Once the park proposal was publicized, political turmoil, initiated by other skeptical indigenous political leaders, ensued and these leaders created relationships with the opposition parties<sup>8</sup>. In 2003, the central government's budget for the

---

<sup>8</sup> Until 2000, the national park authority under the Department of Interior had not been in discussions with indigenous peoples living near or within national parks as land was controlled by National Park Law. Three national parks affect the lives of indigenous peoples: Yu-Shan National Park, Shei-Pa National Park, and Toroko National Park. The indigenous peoples impacted are the Atayal, Bunun, Tsou, and Toroko tribes. In some senses, the National Park Authority embodies a great threat to the traditional practices of indigenous peoples (Sung 1999). Thus, the Maqaw National Park proposal has provoked some indigenous politicians even though the “new” park contains goodwill toward indigenous peoples as promised by the New Partnership Policy. On the Maqaw issue, indigenous opinion leaders are split on the new park. The counter-camp allied with the opposition party and

new national park was frozen by the Legislative Yuan<sup>9</sup>, which demanded increased communication with local people and input of local knowledge, such as knowledge about traditional territories and traditional ecological knowledge. The concept of “co-management” as a new concept remains only a slogan and has not been implemented.

Importantly, the proposed co-managed national park demands that the active involvement of local indigenous people be increased. Moreover, the input of Atayal traditional ecological knowledge is regarded as an urgent need. The reason for the need is based on discussion of the co-management mechanism of the ninth round meeting held by the Maqaw National Park Advisory Committee in April 2002. The official management document stated that the national park should “assist neighboring indigenous villages in developing autonomous eco-industries, tourism, services and research facilities. Furthermore, through exploration of traditional ecological knowledge, the national park can enhance nature appreciation and protect cultural heritage.” The document also states that the national park should “respect the traditional territory of indigenous people and their living space.” However, it should be noted that at that time no official records of traditional ecological knowledge and traditional territory in the Maqaw area existed. Most importantly, such knowledge cannot be understood when not expressed in the Atayal language. Furthermore, this knowledge is held by the local and older generation to whom politicians and academics on the advisory committee have little access. In terms of nature conservation and autonomy of indigenous people, the committee’s conclusions regarding traditional ecological knowledge and territory are far from the real social context of the Atayal people.

Additionally, although indigenous ecological knowledge has become gradually recognized by nature conservationists and governmental agencies protecting natural areas, it is frequently seen as promoting stereotypical and romantic perspectives. Rapid social changes of indigenous communities caused by modernization and free-market mechanisms are often neglected. Hence, social and economic problems faced by indigenous communities are also ignored. The impact of cultivated economic plants on the environment is largely negative. For instance, roads built for transport of agricultural products to markets are often inappropriate. As for the huge amount of forest areas and forest resources still largely not harvested, most of which are situated in government-owned forests and protected areas, indigenous people are forbidden to use these areas under existing laws. As a whole, this situation is far from the romanticized discourse of indigenous people as guardians of mountains and forests (Lin 2007). This study does not analyze the capabilities of indigenous people to manage forests, but rather indicates the existing social conditions, such as market economics, the collapse of tribal communities, population outflow and existing laws and regulations. Notably, these social factors play important roles in endangering indigenous languages simply because indigenous people are rapidly alienated from their environments and related language practices. Classic studies on social change impact on indigenous peoples’ culture in Taiwan can also be found in Huang (1973, 1993).

adopted a strong oppositional strategy by holding several protests in front of government offices, including the presidential hall, that pressured the government to cut the budget for the new national park (Lee 2004).

<sup>9</sup> Taiwanese parliament.

The difficulty in identifying traditional ecological knowledge in the proposed national park relates to a decrease of spoken Atayal in people's daily life and a lack of understanding of the relationship between Atayal language and the Atayal people's living environments. Without noting the real context, any committee's decision related to traditional ecological knowledge and traditional territory will be based on the cultural prejudices of committee members and not on Atayal cultural practices. Despite the existence of linguistic studies and Atayal language professionals, no linguists were used by the committee as consultants at that time (Li, 1980, 1981, 1982, 1985, 1995; Huang 1993, 1995, 2002; Rau 2004; Chang *et al.* 2005; Huang 2006a, 2006b). The Council of Indigenous Peoples has devoted substantial resources to protect indigenous languages; however, teaching and learning activities are limited to passing a certification examination rather than learning about the history, geography, culture, and social situation of indigenous peoples by thinking in a native tongue (Wu 1999; Sun 2005). However, the real need for native language from the perspectives of the proposed Maqaw national park and the "New Partnership Policy" resemble two lines that run parallel and never cross. Unfortunately, the proposed development of the Maqaw national park was stopped in 2003 by the Legislative Yuan primarily due to a strong distrust of the process by indigenous political leaders and lack of support from local indigenous communities. National parks have had a considerable impact on indigenous people living in mountains. Because of protected areas such as national parks, indigenous peoples have no access to lands reserved by the state. Governmental restrictions on access to traditional lands hinder the ability of indigenous peoples to acquire traditional knowledge. Subsequently, much vocabulary related to traditional practices, such as hunting and gathering, has been lost. Moreover, language related to traditional skills, rituals, taboos, and even traditional beliefs associated with the land are endangered. The proposed national park brings about an opportunity to revitalize the language practiced in a local environment; however, due to a lack of interaction with linguistic professionals and the omission of a socially-engaging indigenous language policy, the proposal failed.

To certain extent, the Maqaw national park embodies an encounter between the ideals of nature conservation and indigenous traditional ecological knowledge. The ideals of nature conservationists can be expressed easily in Mandarin, the dominant language in Taiwan; however, the expression of traditional indigenous ecological knowledge is codified in indigenous languages and daily practices. The discussion of indigenous ecological knowledge is closely related to how people understand nature and their relations to nature.

The question, "What is nature?" has long been discussed in many different academic disciplines. Glacken (1967) and Worster (1979) traced the transformation of the concept of "nature" in the western societies. Their research revealed that understanding "nature" has social and historical aspects. That is, "nature" is not naturally presented as it is, rather it is presented in the context of imagined characteristics derived from different social or cultural contexts. The question of who interprets these imagined characteristics is of primary importance. Wright (1992) argued that we must give up viewing "nature" as external and objective. However, by doing so, one need not give up the possibility of pursuing reasonable knowledge. Furthermore, Wright suggests that a "wild" knowledge exists in which languages are the key media between humans and the world they inhabit. In the Maqaw case, this project found that understanding the Atayal language is critical for constructing indigenous ecological knowledge and dialogue with nature conservationists and the state.



**3.2 CASE II: LOST IN TRANSLATION—THE NATIONWIDE INDIGENOUS TRADITIONAL TERRITORY SURVEY.** A map can be understood as “a conventionalized image representing selected geographical features or characteristics designed for use when spatial relationships are of primary relevance” (Board 1991). Maps have long been used by indigenous people in diverse ways; for example, vocal chanting is a common method by which the Atayal and many other indigenous peoples in Taiwan communicate how ancestors moved from one place to another. Indigenous ways of mapping were marginalized in the arena of spatial politics by the progress of “modern science” driven by an assumption of universality and objectivity intertwined with the authority of state power since the 16<sup>th</sup> century (Escolar 1997; Harley 2001; Turnbull 2003).

Motivations to integrate indigenous people into the mapping process first came about through the simple academic interests of anthropologists conducting ethnographic research in North America (Boas 1934; Boas 1964). During the 1960s, Boas and his students conducted a series of investigations in North America and worked with indigenous groups to map their daily activities and ecological practices (Natcher 2001; Chapin and Lamb *et al.* 2005). In these studies, indigenous people were more likely “being mapped” than “mapping.” A breakthrough event was the success of the Cree people in applying a map generated with an anthropologist to claim their land rights and reject a planned hydroelectric plant planned for James Bay in the 1970s (Natcher 2001). Their success demonstrated the possibility for indigenous people to construct maps according to their interests, particularly land rights. Diverse methods emphasizing different aspects of mapping indigenous lands have been developed worldwide and have resulted in diverse terminologies (e.g. Land Use and Occupancy Studies (Usher 1990; Usher *et al.* 1992) in Canada; “participatory mapping” (Fox 1998); “community mapping” (Bennagen and Royo 2000; Eghenter 2000; Fox 2002) in Southeast Asia; and Participatory Geography Information System (Seiber 2000) in the United States). Nevertheless, such mapping represents “a shift in the way cartography is undertaken and used” (Chapin and Lamb *et al.* 2005) that “increases the power of people living in the mapped areas to control representations of themselves and claim to resources” (Peluso 1995). The participation of indigenous community members in mapping processes is a common method of indigenous self-representation, and a way to empower indigenous communities.

In 2002, the Taiwanese government launched a nationwide Indigenous Traditional Territory Survey (ITTS) to fulfill President Chen’s commitment in his “New Partnership Policy.” In this survey, community maps, community participation and computer-based GIS were integrated to identify the territories and traditional knowledge of indigenous communities. This survey was a response to the growing assertion by indigenous groups for inherent land rights. However, without sufficient awareness of epistemological differences between languages, the opportunity for indigenous self-representation was lost.

**3.2.1 MISTRANSLATED “COMMUNITY MAPPING”.** After struggles over the constitution and the central governing party in the ’80s and ’90s, the indigenous movement in Taiwan turned to local and place-based issues. For instance, the Rukai people in the Hau-Cha village organized and successfully resisted a governmental project that planned to build a reservoir downstream on the Ai-Liao River that would require moving the Rukai people and submerge their heritage permanently. In Danayiku, the Tsou people organized

to protect their fishery and village. By using the clan system in patrolling the Danayiku River, villagers successfully eliminated unlimited tourist fishing activities, which were considered as the primary cause of the ruined fishery. The above case on co-managing the Maqaw national park in the Atayal area is another example demonstrating that the focus of the indigenous movement has shifted from the constitution and central government concerns to peoples and the places in which they live.

Two trends can be identified along with this shift. First, “Bu-Luo” (部落) has become a popular term and appears in many discussions of issues related to indigenous peoples. “Bu-Luo”, initially a Mandarin term used by anthropologists, refers to “tribal settlement” in non-western and underdeveloped tribal societies. However, as the indigenous movement shifted its concerns, “Bu-Luo-ism” (部落主義)— regardless of its definition of tribalism in anthropology — was utilized by indigenous activists to highlight a new movement strategy that emphasizes grass-roots power and seeks local knowledge. In fact, the term “Bu-Luo” has become generally synonymous with the indigenous communities, even though the way in which a “community” is organized varies with different peoples and different regions.

Second, GIS was first adopted by indigenous peoples to make sense of their “place”. Even though some zoologists have conducted studies using GIS and worked corporately with indigenous people, Taiban Sasala, a Rukai tribal intellectual and activist (formerly the director of Kaosiung City’s Bureau of Indigenous Peoples) was the first person to introduce the concept of “community mapping” in Taiwan. Taiban was also the first to employ modern GIS to map the Rukai’s traditional territory and organize a team, including village members, to survey local names and natural resources in these areas. During controversies associated with the co-management scheme for the proposed Maqaw national park, Taiban was the first to publicly discuss the importance of community mapping. In a letter to the editor published in China Times, a national newspaper, he advocated “before the setting of Maqaw national park, a “Bu-Luo-Di-Tu” (部落地圖) should be made in advance.” This letter garnered many positive responses in Taiwan.

“Di-Tu” (地圖), is the noun for “map” in Mandarin. “Bu-Luo Di-Tu” (部落地圖) refers to “a map of a tribal settlement.” Although community participation in the mapping process is acknowledged as important in “community mapping” discussions in academia, this translation can very easily mislead readers, and be misunderstood as emphasizing the outcome of mapping—that is, a map of an indigenous community.

**3.2.2 MISLED SURVEY.** In the ITTS launched in 2002, the Council of Indigenous Peoples, the financial sponsor of this project, announced its goal of mapping all indigenous communities (more than 600 communities located in 55 indigenous townships) over a three-year period. Through open tender, the Council of Indigenous Peoples contracted a research team of mainly geographers from different universities to conduct surveys starting in 2002. In the first year of the survey, 30 indigenous communities were chosen as exemplar locations for the survey. Among these communities, the Atayal community of *Smangus* was chosen as a model to demonstrate official mapping practices. Research team members were assigned to facilitate community mapping. Although “community” was in the project name, the project goal was made without first gaining community input. The 30 communities in the first year were chosen based on geographers’ familiarity with the

communities. Furthermore, the project degraded community participation to a very low level—this is discussed further after the following description of the mapping procedure in the second year.

In the second year, a relatively larger team was organized and the survey area was extended to all 55 indigenous townships. The mapping work was mainly conducted using the following procedure (Chang 2002; Chang 2003; Chang 2004):

**(a) Organizing Working Groups and Accommodating GIS Sets.** The mapping procedure began by organizing working groups for surveying and generating the GIS sets. The survey working groups organized into three levels. In the first level, geographers and other professional researchers comprised the research team in charge of the following tasks: generating the GIS sets; organizing workshops (which is explained further in the next section); facilitating mapping work in the other two levels (basically, each facilitator in the research team was assigned to map 3–5 townships); collecting paper-based maps and records of oral histories from local township functionaries; compiling these records and histories into digital data; and presenting these data as digital maps. Second, functionaries from 55 indigenous township governments served as facilitators at the township level, and were in charge of collecting the paper-based maps and records of oral histories from community mapmakers. Financial funding for this mapping work in communities came from the Council of Indigenous Peoples and was distributed through township governments. In the third level, some community members were invited to join the mapping work as community mapmakers. These community members provided oral histories and identified locations, boundaries, and areas on paper-based contour maps according to their knowledge or information from community members.

The GIS settings included preparing the GIS software, digital contour maps, satellite imagery and aerial photographs, providing GIS software to township governments and providing the 1:25000 paper-based contour maps to community mapmakers.

**(b) Plenary Session and Regional Workshops.** A plenary session was held in advance of field visits. All township-level facilitators and community mapmakers were invited to attend the sessions. The research teams, survey goals, concepts of participatory mapping, operational processes and GIS settings used in the following workshops were introduced. Regional workshops were then held separately in or between communities according to local needs. Community mapmakers joined the workshops and were taught basic skills of ethnographic information recording and contour map reading. Satellite imagery, aerial photos and 3-D maps in GIS were shown to local mapmakers who used them to identify targets on paper-based contour maps.

**(c) Information Coding, Compiling and Public Hearings.** The oral histories included community events and their relative locations. Boundaries and areas were recorded in indigenous languages and translated into Mandarin by community mapmakers, and identified on the paper-based contour maps collected by township facilitators. The data collected were transferred to research teams and compiled in a text-based report, atlas and GIS database. Research team facilitators attended public hearings held by township facilitators in township governments. With the atlas primarily compiled through GIS, the research team

facilitators and township facilitators jointly explained and interpreted the data collected about community residences and required the residents to verify the data. After public hearings were held in each township, the research team further revised the text-based report, atlas and GIS database. The revised text-based report and atlas were presented to the Council of Indigenous Peoples and then published.

In the survey's third year, a similar procedure was executed to complete the survey. By the end of the third year, approximately 464 indigenous communities belonging to 12 different tribes located in 55 townships were mapped. About 3700 native place names in indigenous languages (translated into Mandarin) were recorded along with folk stories, myths and oral tales attached. Some communities have well-defined territory boundaries or boundaries of hunting/cultivating territories (Chang 2004).

Although the initial objectives of this survey were met, community participation was decreased to a minimum. Community members were only used for collecting information and operating computers. Indigenous communities did not participate in the debate over what constitutes a "territory" in indigenous epistemologies. Informed by the term "Bu-Luo Di -Tu," the survey sponsor focused excessively on the "map" and the outcome of mapping geographical boundaries of "community" (which is the necessary geometrical element forming a territory in a modern state), and ignored the dynamic meaning of "community participation" in the mapping process. In the next section, we discuss how insufficient indigenous participation resulted in mistranslation of the indigenous concept of "territory".

**3.3.3 MISTRANSLATED "TERRITORY."** In Mandarin, the term "Ling-Yu" (領域), which means "territory" in English, has been utilized by indigenous movements prior to the 2002 ITTS. "Ling-Yu" (領域) was typically illustrated in the context of sovereignty claims against the state, rather than as a well-defined geographical boundary of an individual community. Dr. Ming-Hui Wang, for instance, a member of the Tsou tribe, launched the indigenous autonomy movement to claim traditional indigenous territories. Dr. Wang is a human geographer teaching at a university. Over 20 years ago, Wang first began mapping his own tribe's traditional territory and used these materials to write his master's thesis (Wang 1989). The Tsou then had a chance to create their own tribal council, which was the first among all tribes in Taiwan and independent from official governments. Dr. Wang was a key figure in the establishment of tribal councils. Due to his academic career, Dr. Wang can easily access rich resources and shape public opinion about indigenous land issues.

Masa Towhu is an Atayal elder who is experienced in dealing with different colonial powers, namely, the Japanese and KMT. He has devoted over 30 years to fighting for Atayal traditional territory. Masa was trained by the Japanese to use modern maps made by the Japanese. However, he utilized these official maps from different colonial powers to expose their different attitudes toward indigenous lands, and further to publicize that Atayal land had been stolen. As the second largest tribe in Taiwan, the Atayal people are distributed widely from central to northern Taiwan. Elder Masa utilized different maps to document geographical evidence of Atayal villages, even those in ruins. In his relentless pursuit to identify Atayal traditional territory, he recorded rich oral histories from different villages. Additionally, he is also involved in a law-suit and initiated a social movement to fight for traditional indigenous territories (Lin and Hsiao 2002)

In the ITTS “territory” was detached from its social context and utilized as a goal of mapping the “Bu-Luo Chuan-Tung Ling-Yu” (部落傳統領域), the traditional territories of indigenous communities. The idea that each territory has a fixed geographical boundary is a concept of the modern state and was applied in this project without considering the indigenous notion of territory. Notably, no indigenous language has a word that completely corresponds to “territory.” Furthermore, our finding from studying Atayal traditional hunting grounds indicates a phenomenon of sharing territory. For some indigenous communities, there is no fixed geographical boundary between them. Concisely speaking, the geographical boundary is very much determined by the social relations between indigenous communities.

**3.3.4 MISLED BOUNDARY DELINEATION.** Empirical survey experiences indicated that delineating the boundaries of some indigenous communities is extremely difficult as they change continually over time and because definitions of different ethnic groups differ significantly. As the research team acknowledged and noted in its report, “some boundaries are ambiguous, some territories between communities overlap and some territory maps were confined within the boundaries of existing administrative districts. It is difficult to identify boundaries based on the perspectives of different ethnic groups” (Chang 2004). However, the goal of the Council of Indigenous Peoples was to map the territory of each individual community with concrete boundaries. Thus, this goal also urged the survey team to identify concrete geographic boundaries.

In Atayal language, different terms refer to various social-spatial relations. *Gaga*, for instance, refers to a set of customs, rules and rituals driven from *Utux* (the highest spirit) belief. Meanwhile, this term also refers to a group of individuals who follow the same set of customs, rules and rituals. *Qalang* refers to the residence of a group of individuals, similar to the definition for “settlement” in English. The relationship between *Gaga* and *Qalang* varies across regions. In some regions, one *Qalang* is home to one *Gaga*, while, in others, one *Qalang* may have many *Gagas*. Conversely, many *Qalangs* can belong to one single *Gaga*. *Qyunam* may be the Atayal term closest to the term “territory” in English. A lineage group normally shares a *Qyunam*, which typically occupies a watershed for purposes of hunting, farming and fishing. Although easily deemed as “territory”, *Qyunam* differs somewhat from the concept of territory in modern societies.

As for the *Smangus*, which is one *Qalang* of the *Mrqwang* lineage, it shares a *Qyunam* with all other *Qalang* in this lineage. Inside the *Qyunam* of *Mrqwang* lineage, each *Qalang* acknowledges its responsibility to *Malahang* which, in the Atayal language, refers to “taking care of” their *Qyunam*. However, when individuals from *Mknazi* (another lineage group that occupies a nearby watershed) came to hunt in the *Qyunam* of *Mrqwang*, they were welcomed and even given more prey because, in the *Mrqwang* elder’s word, “*Mknazi* live farther and have more difficulty in capturing prey”. *Mrqwang* people were also welcomed and given more fish when fishing in the *Qyunam* of *Knazi*. Rather than a total, exclusive ownership, the relation between people and their *Qyunam* is flexible and contingently determined by social relations.

At the 2002 ITTS, *Smangus* was the only *Qalang* selected in the *Mrqwang* lineage. When *Smangus* community members were asked to identify their “territory”, they identified the whole area of the *Qyunam* of *Mrqwang* lineage. The published report of 2002

survey designated this area as the “traditional territory of *Smangus*”, subsequently raising tensions between *Smangus* and other *Qalang* of *Mrqwang* lineage. In subsequent years, after all other *Qalang* of *Mrqwang* lineage groups were included in the survey, this area was re-designated as “the traditional territory of *Mrqwang* group”. In 2007, when the Council of Indigenous Peoples attempted to officially announce the *Mrqwang* Traditional Territory, which will be followed by a new regulation allowing *Mrqwang* people to gather certain natural resources in this area, neighboring *Mknazi* community members angrily and fiercely resisted. The gathering activities have been stringently forbidden by the Forestry Bureaucracy since Taiwan gained independence from Japan in 1945. This new regulation will provide some access, although extremely limited, for the *Mrqwang* communities to utilize the forest. However, for the *Mknazi* lineage, designating this area as “the traditional territory of *Mrqwang* group” implies that *Mknazi* are officially excluded from legal access to this area. However, the Council of Indigenous Peoples was confused over whether the survey team had delineated a “correct” boundary between *Mrqwang* and *Mknazi*. Nevertheless, the fundamental error is that the survey carelessly deemed the Atayal *Qyunam* and “territory” as the same entity without seriously considering language concerns.

**4. LOCAL PRACTICE OF REVITALIZING THE ATAYAL LANGUAGE— A CASE FROM THE SMANGUS VILLAGE.** In this section, an Atayal village, *Smangus*, is utilized to illustrate the local practice of revitalizing the Atayal language. A university-community collaborative course developed by the authors is introduced to show how local Atayal can participate in revitalizing the Atayal language in their local environment. In this way, this work demonstrates the critical link between practical socio-cultural concerns of indigenous communities and their indigenous language practices. Furthermore, the course establishes a forum for speaking an indigenous language in the modern context of ecological education and provides a channel for communication with nature conservationists. The course, entitled “Peoples and the Environment,” is a liberal arts course taught at Providence University. By focusing on the subject of Indigenous Peoples vs. Natural Resource Management, the course enables students to interact with indigenous villages via PAR and Atayal villagers *vice versa*. Theoretically, this course is an ecological education course that considers culture and nature as inextricably linked. This view is in contrast to the nature-culture dichotomy and has earned considerable support by sociologists and geographers in recent years (Soper 1995; Cronon 1996; Macnaghten and Urry 1998; Castree and Braun 2001). Similarly in the field of environmental education, Educating/Learning with Environments is proposed by Gough (1987). Environments here imply multicultural perspectives. In other words, an environment is a product of social construction by actors under various social-cultural contexts. In the process of environmental education, the idea of “environments” should encompass more than the physical, and exceed the mere physical dimension of human society. Consequently, environmental education should focus on groups of people and their interaction with their living environment, the society to which they belong and the culture in which they exist. Therefore, the Atayal culture and language is the core of this course.

The course begins by examining the forestry issue of proposed Maqaw national park, followed by students’ real experience in interacting with nature and culture in the *Smangus* village. The *Smangus* village is a unique Atayal village that has strong interests in issues

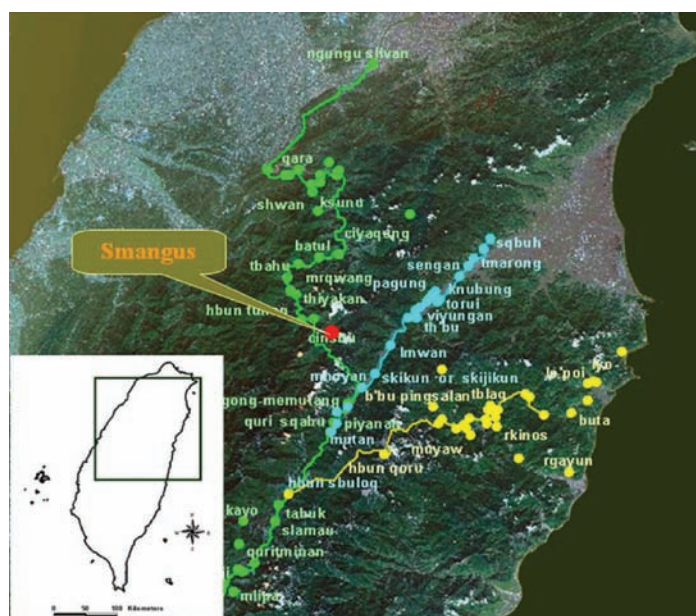


FIGURE 2. The Location of the *Smangus* village from the Atayal Historical Migratory Map

associated with the proposed Maqaw national park and the traditional territory survey. Figure 2 shows the location of *Smangus* from the Atayal historical migratory map. First, the Chaemacypris forest, which plays an important role in the controversy surrounding the national park, is partially overlapped with *Smangus* traditional territory. Before the co-management concept is raised via the government's "New Partnership Policy", knowledge of the Chaemacypris forest was confined to western biological research carried out by the national park authority and Forestry Bureau. Indigenous knowledge of the forest was simply ignored by the outside world even though the forest is alive in daily practices and talked about in indigenous languages. The controversy surrounding the national park and the protection of ancient giant tress by the *Smangus* village has garnered attention from the outside world. This community was then chosen as a model community for the community mapping project initiated in 2003. However, both the proposed national park and the mapping projects have failed to recognize the importance of indigenous language practices and their meaning in daily life. To certain extent, *Smangus* villagers' participation was limited by being only a symbol of a national plan.

The course design allows students to learn through reflection on cultural and environmental values in the abstract and through learning environments outside classroom that feature an interpretation system with villagers in their mother tongue, coupled with consecutive translation into Mandarin. We believe this framework will make it possible to combine local people speaking their native language with knowledge of the local environment. Hopefully, the villagers will develop skills for interpreting the environment for outsiders. Conversely, students of Providence University are also a main course target

and attend classes that are not germane to villagers concerns, such as watching ecological documentaries that enhance student reflection on the human-nature paradigm. Moreover, the course introduction in the very beginning also concentrates on essential material like issues of Maqaw national park and community mapping for introducing students to the Atayal understanding of nature.

To allow students to experience nature in a village, students are taken to the *Smangus* village and nearby forest trails for a two-day visit on one weekend. This part of the course is the foundation for our long-term involvement in PAR on community development and natural resource management over the last seven years. Through exhaustive discussions with villagers, an outdoor course was designed autonomously by villagers according to their understanding of their ecological environment. This course is a manifestation of the social development approach from the perspective of community empowerment. Therefore, when students partake in the two-day village program, they are imperceptibly involved in community empowerment work. Furthermore, students are educated by indigenous people.

The course curriculum is the result of interaction between the teacher, villagers, teaching assistants and students, especially when teaching in the approach of PAR. Despite the fact that the teacher has a crucial role in this course and in conventional instruction, the teaching differs in certain ways. The coordination of workload between the teacher and teaching assistants is a novel approach: in addition to using teaching assistants from the institute, a few assistants from the village were recruited. Research and social practices in the last few years has garnered full assistance and support from the village. Since a student visit to the indigenous village must be arranged, it is necessary that villagers participate, thereby playing an important role in the communication with the village. After all, the village visit requires the most effort — taking a large class into the mountains is a difficult trip to organize. The teaching assistants usually spend 75% of their time coordinating with the villagers. These assistants are partners the authors have worked with for a long time. This fact implies that the relationship between the authors and village assistants is built on solid foundation of trust and cooperation. The outcome of this relationship is rich, not only for students who learn from their experience but also for the host villagers who learn how to arrange a two-day ecological visit for student tourists. After the visit, the host village reviews the activities and opens a space in which they consider future industrial development for the village. Villager participation initiates a social practice in this program. Walking in the forest is in itself an enjoyable activity, and it is an excellent opportunity for students to gain experience in village life. For instance, students observe cultural “performances,” such as staying in a local person’s house, eating Atayal food, and drinking fresh cabbage soup. Through village life, students obtain knowledge of political-economic issues associated with marketing agricultural products and resource exploitation. The predominant goal of these practices is to generate interaction between students and villagers. This teaching approach aims to create and understand the “socio-cultural dimension of nature.” This so-called humanistic dimension comprises the act of generating meaning and value. If a student learns stories about the mountain on which he or she is staying, this mountain is no longer solely a physical entity, it also takes on cultural meaning.



In this course, cultural experience is combined with an ecological program. Students are taken to the indigenous village, allowing them to gain knowledge of the land from multiple socio-cultural dimension of nature. From villager viewpoint, the Atayal language is a basic instrument in which village assistants must be fluent and capable of discussing and organizing course preparation, ranging from coordination, course design, division of labor, and internal training, to interacting with students. These processes combine village lifestyles—the shift between agricultural production and tourism—and experience of village life to create a teaching experience that best shows the vitality of local cultural ecology and helps students consider the question “what is nature?”. The Atayal language has gained an irreplaceable position in the course. We now discuss some examples that demonstrate how the course works and the indigenous language is practiced.

It is worth noting that most of the cultural and ecological instruction by the villagers will begin in Atayal language, followed by translation into Mandarin. During group learning, village assistants teach students traditional place, animal and plant names, mythologies, village history, and about the relationships with each village in that area. Interpretation is, in a way, a representation of local environmental knowledge, which is presented as a narrative by village elders, village assistants sharing their life experiences and the spoken Atayal language, all of which are important to modern Atayal culture. We believe these experiences can only be gained in an indigenous village.

***Sbalay: the cleansing ceremony for the start of the two-day trip.*** When the course participants arrive at the village of the first day, the chief of the *Smanugs* village briefly outlines his expectations of the students during this two-day village program. *Sbalay*, the cleansing ceremony through the act of watering two saplings, serves as a blessing for each student. Afterwards, each student introduces himself or herself. In line with Atayal tradition, village elders hold the ceremony and speak to Atayal children returning from the world outside the village. Similarly in the course, the elders hope to embrace the souls of the students in the village air—this is a very important Atayal ceremony. The hope is that students can “feel” the *Sbalay* via the act followed by translation of a local interpreter. The meaning of *Sbalay* is thus introduced.

***Pnhwan: fire place.*** Those who grow up in the village generally have many memories of the elders, especially of tribal stories told endlessly at the Pnhwan (fire place). On the first night, students are introduced to the fire place custom, sometimes accompanied by a barbeque and the making of millet cakes. Normally, a comprehensive history of the village and activities are told on this occasion within which two other main activities related to Atayal traditional knowledge and language are introduced: (1) the introduction of the whole process of millet growing through a slide-show, often followed by a DIY activity of making millet cake; (2) the half-day walk in Koraw eco-park forest trail and introduction of ethno-biological knowledge on the next day. In the following, the main teaching and learning materials are provided.

#### 4.1 THE PROCESS OF MILLET GROWING AND PROCESSING.

**Trakis: millet.** The story of *Trakis* (millet) is always the first story told by elders. This story starts with the complete production process from seeding, harvest to manufacturing food products and focuses on *Sm'atu* (the millet seeding ceremony), which is observed with many taboos and symbols that signal the beginning of a new year. However, students cannot participate in all stages of the process; thus, some stages are discussed during a slide show on the first night. Additionally, teaching material is prepared in advance by recording the oral histories. The following knowledge in Atayal language is explained in the course. Moreover, photos are provided for better understanding of how the course program works.

**The sign for starting growing millet.** Tribes and villages vary; take the upstream village of *Llyung Papak* (the watershed of *Papak*) as an example. Village inhabitants usually use the time at which the cherry trees in *Tanan Sayun* (a place name) blossom as the time for the millet seeding ceremony. Figure 3 shows that the sight of red and white buds signals the start of the millet-seeding season, which ends when the cherry blossoms fall. According to the elders, if seeded after the cherry blossoms fall, the millet harvest would be poor due to fierce storms.



FIGURE 3. Cherry Blossom as an Indicator of Beginning of Growing Millet

**Sm'atu: the millet seeding ceremony.** First, *sm'atu* signals the start of work for a new year, and literally means the act of spreading seeds and covering them with soil. Before seeding millet and corn, villagers hold a ceremony to pray for a good harvest. Second, the term officially refers to the beginning of a new year for the village. Figure 4 shows a chicken is killed and offered to God in heaven and ancestors. After the chicken is killed, its blood is dripped into a bamboo basket and on the seeds, hoes, and in the fields to pray for a good year and a *qoyat* (blessing) that is as abundant as the chicken blood. The most important aspect of killing the chicken is *qmes* (avoiding disaster); that is, pray for a peaceful year and plentiful harvest, without disturbing the evil spirits.



FIGURE 4. The Essential Elements for the Millet Ritual

***Tmubux*: seeding.** Whether the fields are seeded with millet on the next day is determined by a dream divination on the previous evening. Figure 5 shows villagers are seeding in the field. A dream of a river suggests a great millet harvest, whereas a bad dream tells some to seed on another day.



FIGURE 5. Millet Seeding

***Lmahing: weeding***

The elder says: Hbku te magal qu abaw nya qu Trakis lga, 'san ta mahing ru bkgun qu yaya nya ga helaw mrkyas hopa ru bwaxun.

“When each millet plant grows five leaves, it is time to weed. Weeding helps make room for the plants to grow strong and tall in an adequate space and brings an abundant harvest.”



FIGURE 6. Weeding

***Smi ukuw qoliy: setting up rat traps***

The elder says: Babaw nya hengan lga, aring zik pinturing ru rmapit (kugan) beh sesaw na qmayah. Ini wahi kmyut na Bhut 、Tuku 、Qoliy ru nnanu.

“When the millet begins to tassel, we know the reaping season is coming soon. To prevent animals such as Red-bellied Squirrels, the Formosan Striped Squirrel, Spinou Country Rat, and others from eating the millet, we set up traps around the millet fields.”



FIGURE 7. The Trap to Catch Rats That Might Eat Millet

***Smi tlpak mlawa: setting up devices that repel birds***

The elder says: Mrrang (aring mhebung qu bway trakis) lga. 'san smi tlpak ru ki'an mlawa kryax. aki ini wahi maniq na Qayu ru Pzit.

“Before the millet tassels, we need to set up devices that repel birds around the fields and have people guard the fields so that the birds will not come and eat our crops.”



FIGURE 8. The Device to Repel Birds

***Kmluox: harvest***

The elder says: Tehoq qu ryax kmluox lga, cingay qu gaga nya .

Ini p'sang.

Ini pskura gleng t'asiy.

Ini kayal mha moyay.

Ini piyu atu' kmluox.

Ini kayal mha usuw qu bengan.

kmayal qu bnkis Atayal:

Ini ta glgiy lga, baq ini bhoywaw ru cipoq qu knluox ta qutux kawas la. cingay na qu gaga nya ga, nyux ini bruy na.

“There are many regulations (Gaga) for the time of millet harvest, which are norms and taboos associated with millet production. For instance, noise, inconsistent harvesting direction, coughing forward, cries of hunger, and calling the millet heavy are not allowed. These taboos are passed down from the ancestors. If we don't harvest in obedience with the Gaga, we will end up with a harvest of millet that is less than enough.”



FIGURE 9. Millet Harvesting

***Pgyan: drying the millets in the sun***

The elder says: Bla<sup>q</sup> kayay mu qu wagi, pgyan bengan cyugal ryax ru syukun, pgyan bway cyugal ryax. kyay lga psktan.

“When the sun shines on the land, the harvested millet should be dried under the sun. Both sides of each tassel need to be dried for three days. Until millet are thoroughly dried, each tassel has to be rearranged.”



FIGURE 10: Millet Drying

***Tt'an bengan*: cleaning out the tassels**

The elder says: Galun pucing (soki ˘ lalaw) tt'an qu ami, cint'an ami qu qaya nya ga, ini pqli baq stngahoq.

“After that, we cut the superfluous stalks that cannot be trodden upon. If you tread upon them, you will get a malignant boil.”



FIGURE 11. Millet Cleaning

***Skun khu'*: storing millet in the barn**

The elder says: 'son smi khu' ru blequn stluhung (cinlhongan) sa ska khu'.

“After drying and cleaning the millet, we store the stalks in bundles in the barn, and will call the production work finished.”



FIGURE 12: Millet Storage



**4.2 THE WALK IN KORAW ECO-PARK FOREST TRAIL.** The activity includes:  
 FIGURE 13: Students' Reflection Time in the Forest

the ethno-ecological interpretation along the crest line and the trail, the story of the *lhyux na bnkis* (ancestral cave), traditional hunting culture such as demonstrating how to set up traps, use fire and cook food. Finally, Figure 13 shows that a group discussion in the forest is a very critical time for reflection of students and the villagers.

**4.3 REFLECTION TIME INVOLVES GROUP DISCUSSIONS WITH THE PARTICIPATION OF VILLAGERS** Group discussions are enjoyable and reflective, and often held in the forest. Before the two-day program ends, each group discusses the feelings and experience before and after the village visit as feedback to instructors and villagers. To encourage the students, the lecturers and teaching assistants from the village also share how they feel about the interaction during the two days. Figure 14 shows the lectures of villagers in the forest.



FIGURE 14: The Forest Lectures Led by Smangus Villagers



**5. CONCLUSION.** In summary, the examples and cultural practices from *Smangus* are distinct from state indigenous policies that we analyzed above. On the contrary, through autonomous research and investigation, *Smangus* manifests as the relationship between village development and ecological systems in an actual collaborative interpretative program with a university, and, importantly, involves practice of an indigenous language. The difficulties associated with land use under state control are discussed. Until now, the Reserve Policy has confined development to reserve land and encouraged indigenous peoples to grow economically viable crops. This decision resulted in environment degradation and stigmatized indigenous peoples. However, *Smangus* creates a contemporary social system based on traditional knowledge and is extended to a new construction of tribal culture and ecological conservation and industry that are harmonious with nature. These acts not only consider the concepts of modern conservation but also combine elements of industrial development and cultural inheritance. Many ways exist for creating a culture that lives in peace with nature, some of which may be more important than concerns for the environment only. Moreover, the daily life practices of *Smangus* villagers also emphasize that local knowledge is rooted in the place where knowledge is produced, especially when the mother language is prevalent.

Local participation is a recent and new aspect of Taiwanese nature conservation and the indigenous peoples' movement under rapid social change. Local participation highlights the importance of local management of their surrounding environments rather than nature conservation managed by specialists. This concept is built upon the idea of diverse environmental values, which indicate that different communities have different views toward the environment that is most suitable for social development. The idea of local participation also affirms the environmental understanding of local people. In other words, the so-called local knowledge has social value. The knowledge of local people is not less useful than scientific knowledge, even when they differ. This local knowledge is always based on daily life and local cultural, historical, economic and political systems. This knowledge, additionally, is usually expressed through indigenous languages. Thus, the knowledge has considerable accessibility and popularity. People in Taiwan have recently become aware of the importance of integrating nature conservation with the concerns of daily life. Some local environmental protection and indigenous organizations have started participating in local environmental affairs. This trend of local participation suggests that environmental protection must be a lifestyle-based social process linked with cultural politics. This chapter brought this issue into the context of indigenous communities and argued that indigenous language practices are extremely important to local participation and construction of local knowledge. To a certain extent, local knowledge will be difficult to reveal without the daily practices of indigenous languages. Consequently, the insufficient or inadequate understanding of indigenous peoples' local knowledge will also create improper imagination of socio-political policy such as the cases we have examined above. For this reason, this chapter further indicated that a lack of active linguistic input to the government's indigenous socio-political policies — based on the spirit of the “New Partnership Policy” — will be a great loss.

## REFERENCES

- BENNAGEN, PONCIANO, and ANTOINETTE ROYO. 2000. *Mapping the Earth, Mapping Life*. Quezon City, Philippine: Legal Rights and Natural Resources Center.
- BOARD, CHRISTOPHER. 1991. Report of the Working Group on Cartographic Definitions. *ICA Newsletter* 18:14-15.
- BOAS, FRANZ. 1934. *Geographical Names of the Kwakiutl Indians*. New York: AMS Press.
- BOAS, FRANZ. 1964. *The Central Eskimo*. Lincoln: University of Nebraska Press.
- CASTELLANET, CHRISTIAN, and CARL F. JORDAN. 2002. *Participatory Action Research in Natural Resource Management*. New York: Taylor & Francis.
- CASTREE, NOWL, and BRUCE BRAUN. 2001. *Social Nature: Theory, Practice, and Politics*. Oxford: Blackwell.
- CHANG, CHANG-YI. 2002. Report of Indigenous Traditional Territory Survey. Taipei: Council of Indigenous Peoples.
- CHANG, CHANG-YI. 2003. Report of Indigenous Traditional Territory Survey. Taipei: Council of Indigenous Peoples.
- CHANG, CHANG-YI. 2004. Report of Indigenous Traditional Territory Survey. Taipei: Council of Indigenous Peoples.
- CHANG, NAI-HUNG, and YUAN-RUE SHU. 2001. Natural Seeding and Seedling Occurrence in the Chamaecyparis Forest at Chilan Mt. Area. *Taiwan Forestry Science*(16):321-326.
- CHANG, SU-JUEN, and WEN-YU CHIANG. 2005. Language Ability, Language Using and the Continuity of Language :A Case Study of Atayal Silaq Village. *Journal of Language and Literature* 12:139-154.
- CHAPIN, MAC, and ZACHARY LAMB. 2005. Mapping Indigenous Lands. *Annual Review of Anthropology* 34:619-638.
- CHEN, MEI-JU. 2000. Ways of Implementing Taiwan's Aboriginal Language Education from the View of Multicultural Education. *Journal of Education Studies* 45:173-202.
- CRONON, WILLIAM. 1996. *Uncommon Ground*. New York: Norton.
- EGHENTER, CRISTINA. 2000. Mapping people's forests: the role of mapping in planning community-based management of conservation areas. In *Indonesia. Peoples, Forest Reefs Program Discussion Paper Series*. Washington, DC: World Wildlife Fund.
- ESCOLAR, MARCELO. 1997. Exploration, cartography and the modernization of state power. *International Social science Journal* 49:55-75.
- FOX, JEFFERSON. 1998. Mapping the commons: The social context of spatial information technologies. *Common Property Resource Digest* 45:1-4.
- FOX, JEFFERSON. 2002. Siam Mapped and Mapping in Cambodia: Boundaries, Sovereignty, and Indigenous Conceptions of Space. *Society and Natural Resources* 15:65-78.
- GLACKEN, CLARENCE. 1967. *Traces on the Rodian Shore*. Berkeley: University of California Press.
- GOUGH, NOEL. 1987. Learning with environments: towards an ecological paradigm for education. In *Environmental Education: Practice and Possibility*, ed. Ian Robottom, 49-67. Victoria, Australia: Deakin University Press.
- HARLEY, JOHN BRIAN. 2001. *The New Nature of Maps*. Baltimore: John Hopkins University Press.
- HU, AI-RUO. 2002. A Study on Educational Policy over Taiwan Aboriginal Languages. *Journal of Institute of Dr. Sun Yat-Sen's Doctrine* 24:153-173.
- HUANG, HUI-CHUAN. 2006a. Squiliq Atayal syllable onset: Simple or complex? In *STREAMS CONVERGING INTO AN OCEAN: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. Dah-an Ho, 489-505. Taipei: Academia Sinica.

- HUANG, HUI-CHUAN. 2006b. Resolving vowel clusters: A comparison of Isbukun Bunun and Squliq Atayal. *Language and Linguistics* 7(1):1-26.
- HUANG, KAI-JU. 2004. A literature-review study on Transformation of Timber Production Institution in Taiwan, Department of Forestry, National Taiwan University: Mater Thesis.
- HUANG, LILLIAN. 1993a. *A Study of Atayal Syntax*. Taipei: Crane Publishing Co.
- HUANG, LILLIAN. 1995. *A Study of Mayrinax Syntax*. Taipei: Crane Publishing Co.
- HUANG, LILLIAN. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3(2):197-225.
- HUANG, YING-KUEI. 1973. Economic Adaptation and Development in an Aboriginal Settlement in Central Taiwan. *Bulletin of the Ethnology, Academia Sinica* 36:35-55.
- HUANG, YING-KUEI. 1993. Crop, Economy, and Society among the Bunun of Taketonpu. *Bulletin of the Ethnology, Academia Sinica* 75:133-169.
- LI, KENG-CHENG. 2004. The Study on the Promotion of Maqaw Cypress National Park within Civil Society in Taiwan, Department of Ecology, Providence University: Master Thesis.
- LI, PAUL JEN-KUEI. 1980. The phonological rules of Atayal dialects. *BIHP*. 51(2):349-405.
- LI, PAUL JEN-KUEI. 1981. Reconstruction of proto-Atayalic phonology. *BIHP*. 52(2):235-301.
- LI, PAUL JEN-KUEI. 1982. Male and female forms of speech in the Atayalic group. *BIHP*. 53(2):265-304.
- LI, PAUL JEN-KUEI. 1985. Linguistic criteria for classifying Atayalic dialect groups. *BIHP*. 56(4):699-718.
- LI, PAUL JEN-KUEI. 1995. The case-marking system in Mayrinax, Atayal. *BIHP*. 66(1):23-52.
- LIN, YIH-REN, and HUI-CHUNG HSIAO. 2002. Contesting Aboriginal Community Mapping: A Critical View from Local Aboriginal Participation in the Natural Resources Management of Taiwan. Paper presented at *Society for Conservation Biology 16th Annual Meeting*, Canterbury, UK.
- LIN, YIH-REN. 2004. Cultural Construction of Natures: Contesting "Forest" Metaphors of Proposed Maqaw National Park. *Museology Quarterly* 18(2):25-38.
- LIN, YIH-REN. 2005. The relationship between indigenous people and the land: The perspectives of ecology and traditional resources rights. In *The Recordings of Discussing the Proposed Indigenous Chapter of the Constitutional Amendment*, Taipei, Council of Indigenous Peoples.
- LIN, YIH-REN, and KENG-CHENG LEE. 2005. Cypress forest conservation on Taiwan. In *Principles of Conservation Biology*, eds. M. J. Groom, G. K. Meffe and C. R. Carroll, 131-134. Sunderland, Massachusetts: Sinauer Associates.
- LIN, YIH-REN, JUN-MING LAI, and HUEI-CHUNG HSIAO. 2006. Understanding through actions: Reflection on the development and ecological discourse of Pyanan Village in I-Lan. *Journal of I-Lan Documentation* 27:191-242.
- LIN, YIH-REN. 2007. The Chief Seattlization of Nature Conservation in Taiwan: A Radical Perspective on Indigenous Traditional Ecological Knowledge. *Chung Wai Literary Quarterly* 36(3):15-62.
- MACNAGHTEN, PHIL, and JOHN URRY. 1998. *Contested Natures*. London: Sage Publications.
- MEI, KUANG. 1994. Word order, case and theta-agreement in Mayrinax Atayal. Paper presented at *The First Symposium on Austronesian Languages of Taiwan*, Taipei.
- NATCHER, DAVID. 2001. Land use research and the duty to consult: a misrepresentation of the aboriginal landscape. *Land Use Policy* 18(2):113-122.

- NOKAN, WALIS. 2003. Give Me a Magau: Magau, Atayal, State and Park. *Historical Monthly* 183:78-88.
- PELUSO, NANCY. 1995. Whose Woods are These? Counter-Mapping Forest Territories in Kalimantan, Indonesia. *Antipode* 27(4):383-406.
- RAU, DER-HWA V. 2004. Lexical similarity, sound change and intelligibility of Atayalic dialects, *Papers on Austronesian Subgrouping and Dialectology*. Eds. by John Bowden & Nikolaus Himmelmann, *Pacific Linguistics*, Canberra, Australian National University. 37-96.
- SEIBER, RENEE. 2000. Conforming (to) the opposition: The social construction of Geographical Information Systems in Social Movements. *International Journal of Geographical Information Science* 14(8):775-793.
- SOPER, KATE. 1995. *What Is Nature?* Oxford: Blackwell.
- STRINGER, ERNEST. 1996. *Action research: A Handbook for Practitioners*. Thousand Oaks: Sage Publications.
- SUN, TA-CHUAN. 2000. *The Ethnic Construction in Crack: Taiwan Indigenous Languages, Cultures and Politics*. Taipei: UNITAS.
- SUN, TA-CHUAN. 2005. Multiculturalism and Indigenous Language Policy. Paper presented at *Series II of The Conferences of Multiculturalist Thinking for Language Policy: The Vitality and Succession of Taiwan Indigenous Languages*, Taipei.
- SUNG, BIN-MIN. 1999. Responding National Parks' Impact to Taiwan Indigenous Cultures. *Journal of National Park* 9(1):65-80.
- TANG, YI-JIE. 2001. The Word-final Glottal Stop in Squiliq, Department of Linguistics, National Taiwan University: Mater Thesis.
- TURNBULL, DAVID. 2003. *Masons, Tricksters and Cartographers*. London: Routledge.
- USHER, PETER. 1990. Recent and Current Land Use and Occupancy in the Northwest Territories by Chipewyan-Denesutine Bands (Saskatchewan Athabasca Region). Prince Albert, Saskatchewan: Prince Albert Tribal Council.
- USHER, PETER, FRANK TOUGH, and ROBERT GALOIS. 1992. Reclaiming the land: Aboriginal title, treaty rights and land claims in Canada. *Applied Geography*(12):109-132.
- WANG, MING-HUI. 1989. The Spatial Organization of Traditional Tsao Society in A-Li Mountain, Department of Geography, National Taiwan Normal University: Master Thesis.
- WHYTE, WILLIAM. 1991. Introduction. In *Participatory Action Research*, ed. William Whyte, 7-15. Newbury Park, Calif: Sage.
- WORSTER, DONALD. 1979. *Nature's Economy: A History of Ecological Ideas*. Cambridge: Cambridge University Press.
- WRIGHT, WILL. 1992. *Wild Knowledge: Science, Language, and Social Life in a Fragile Environment*. Minneapolis: University of Minnesota.
- WU, TIEN-TAI. 1999. The Promotion of Indigenous Language Teaching in a Perspective of Multiculturalism. *Aboriginal Education Quarterly* 10:49-64.

Yih-Ren Lin  
ylin@ms39.url.com.tw

Lahuy Icy  
lahwy@hotmail.com

Da-Wei Kuan (Daya)  
kuan@hawaii.edu

# 9

## Teaching and Learning an Endangered Austronesian Language in Taiwan

D. Victoria Rau<sup>°</sup>, Hui-Huan Ann Chang<sup>°</sup>, Yin-Sheng Tai<sup>°</sup>, Zhen-Yi Yang<sup>°</sup>,  
Yi-Hui Lin<sup>°</sup>, Chia-Chi Yang<sup>°</sup>, and Maa-Neu Dong<sup>†</sup>

<sup>°</sup>*Providence University, Taiwan*

<sup>†</sup>*National Museum of Natural Science*

This paper provides a case study of the process of endangered language acquisition, which has not been well studied from the viewpoint of applied linguistics. It describes the context of teaching Chinese adult learners in Taiwan an endangered indigenous language, the teachers' pedagogical approaches, the phonological and syntactic acquisition processes the learners were undergoing, and applications to other language documentation and revitalization programs. Both qualitative and quantitative methods were used to address the research questions. This study demonstrates cogently that language is a complex adaptive system. In phonological acquisition, the trill was the most difficult phoneme to learn. Systematic variations for the variables (ŋ) and (s) were found to be constrained by both markedness and interference. Furthermore, learners also tended to interpret Yami orthography based on their knowledge of English. In word order acquisition, learners performed much better than expected, partially because the present tense, coded by the SV word order, is the norm in Yami conversations. However, students still inaccurately associated word order with sentence type rather than with tense distinction. The Yami case provides an integrated model for endangered language documentation, revitalization and pedagogical research, which would be of interest to people working with other languages and the language documentation field in general.

**1. INTRODUCTION**<sup>1</sup>. There are several different approaches to the categorization of language endangerment. According to Ethnologue (Gordon 2005), the total number of languages in the world is 6,912. Of the languages listed, 516 are classified as nearly extinct, as “only a few elderly speakers are still living.” A language might be considered threatened when it is used only by socially isolated old folks, a socially integrated population beyond childbearing age, and when it exists only orally, without literacy (Fishman 1991). Krauss (1992) defines three categories of endangerment: (1) moribund languages: no longer learned by children, (2) endangered languages: still learned by children but not expected to be learned by children within ten years, and (3) safe languages: supported by the state and having a large number of speakers. It is estimated that 90% of the extant oral

---

<sup>1</sup> This paper was presented at the International Conference on Austronesian Endangered Language Documentation, Providence University, Taiwan 6/5-6/2007. The research was partially funded by an NSC grant (NSC95-2411-H-126-007), provided to the first author. We would like to acknowledge the helpful comments by the two anonymous reviewers. All remaining errors are our own.

languages will either be moribund or will have disappeared by the end of the century (Terralingua, <http://www.terralingua.org>). Thus, there is a very limited window of opportunity to document and revitalize those languages.

A number of Austronesian language revitalization projects are related to language teaching and learning. One success story is the establishment of Maori immersion and bilingual schools (Maclagan et al. 2006). Maori communities were specifically built to support Maori language use at home and in social institutions for children. Another case is Hawaiian revitalization, as described by Slaughter (1997), whose success was attributed to parents' positive attitudes and well-established immersion programs.

Some efforts have been made to develop web-based interactive language learning materials for Indonesian and Tagalog (Henry and Zerwekh 2002). Software for teaching these two well-documented Austronesian languages has also been produced, such as Northern Illinois University's "Learning Indonesian on the Internet" (<http://www.seasite.niu.edu/Indonesian/>), Hoven's (2003) MMInteraktif, an Indonesian listening comprehension software package, and McFarland's (2006) CAI program for teaching Filipino. However, e-Learning for endangered Austronesian languages is rare.

As a result of the ELDP grant from the Hans Rausing Endangered Languages Project, Rau, Yang and Dong developed e-Learning (i.e., computer-mediated instruction) materials for Yami, an endangered Austronesian language (2007), proposed a pedagogical model for e-Learning (Rau and Yang to appear) and provided a summative assessment of the impact of their e-Learning materials (Rau and Yang this volume).

Yang (2007), in her formative needs analysis of the Yami e-Learning program, provided feedback to improve on its design. She also surveyed the attitudes of Yami teenagers toward the e-Learning program and found that the third year junior high school students on Orchid Island had a more positive evaluation of the websites than did younger students. She reasoned that this might be due to the pressure they had to pass the Proficiency Tests of Aboriginal Languages (PTAL)<sup>2</sup> so that they could get the 35% bonus points on the Basic Competence Test, a High School Entrance Examination. Since there are limited resources<sup>3</sup> for them to prepare for the test, the Yami websites filled this gap.

Although previous studies have informed us how to build a successful Yami e-Learning program, very little research has focused on the process of teaching and learning Yami. This study investigated the context of teaching an endangered indigenous language to Chinese adult learners in Taiwan, the teachers' pedagogical approaches, the phonological and syntactic acquisition processes the learners were undergoing, and applications to language documentation and revitalization in general. We believe the research presented in this paper, bringing together two distinct but crucial areas, can make a contribution both to second language acquisition research and to language revitalization.

**1.1 CONTEXT.** Yami is a Philippine Batanic language, spoken by less than 4,000 indigenous people on Orchid Island (*Lanyu* in Chinese). According to Rau's (1995) sociolinguistic survey, there was only one village out of six on the island where children still used

2 For a detailed discussion of the PTAL, see Huang (2007).

3 The online indigenous language textbooks for both elementary and junior high school students are available from [http://www.alcd.nccu.edu.tw/index\\_0.html%20](http://www.alcd.nccu.edu.tw/index_0.html%20).

some Yami in daily interaction. Yami has been offered as an elective in elementary school since 1998, but Yami is gradually being replaced by Mandarin Chinese. Chen (1998) has compared the language proficiency, language use and language attitude among three generations of Yami and found a language shift to Mandarin and a decline of Yami language ability as age decreases. Lin (2007) re-examined language use and language ability among Yami teenagers and found that while Yami is still spoken in Iraralay, the other five villages show continuing decline in the use of Yami by teenagers with their parents. However, Yami teenagers were found to have a positive attitude toward Yami language and identity and strongly support any efforts to promote or maintain their language. Be that as it may, most of the teenagers admitted that they prefer speaking Mandarin over Yami.

Rau and Yang (2007) describe a model of biliteracy continua of majority and minority languages in Taiwan and discovered a reversal of the relationship in each of the four continua in comparison with Hornberger's model (2003). The less powerful communities in Taiwan are monolingual in a vernacular whereas the more powerful communities are bi(multi)lingual, with one of them being the official language. The addition of English literacy in the mainstream Taiwanese society tends to follow an L1 model, while the addition of Yami literacy on Orchid Island tends to follow an L2 model. In the acquisition of English in the mainstream Taiwanese society, the content is required to be "contextualized" and "authentic" for communicative needs. On the other hand, the content of online indigenous language textbooks for both elementary and junior high school students is mostly "decontextualized" with direct translation from Chinese. Finally, due to early exposure or in many cases simultaneous exposure to English, the media of biliteracy in the mainstream Taiwanese society have become increasingly similar and convergent in terms of their language structures and scripts respectively, whereas the successive exposure to Yami literacy is viewed as dissimilar and divergent from the Chinese structures and character writing to which they are exposed very early in life.

All the above-mentioned shifts of contextual factors from Yami teenagers' language attitudes, language use, and language proficiency to the disadvantaged end of power relations in the continua model have provided the context for our study.

Although revitalization of the aboriginal cultural and linguistic inheritance of Taiwan is well supported by the Taiwanese governments, central and local, the Yami language is still shifting to Mandarin in its own speech community. However the increasing interest in the indigenous languages and cultures in Taiwan has motivated Chinese students to take such courses in response to the Taiwanese government's campaign of globalization and localization.

**1.2 AIMS AND RESEARCH QUESTIONS.** The goal of this longitudinal study is to provide a description of the process of learning an endangered language by a group of Chinese graduate students in linguistics from 2005-2007, using both qualitative and quantitative methods to answer the following research questions.

1. What is the process of teaching and learning an endangered Austronesian language in Taiwan? What are the teachers' pedagogical approaches?
2. What are the sequences of development and interlanguage variability in Yami phonology? What phonemes are difficult to acquire and why?

3. What are the sequences of development and interlanguage variability in Yami syntax? How is word order in Yami acquired by Chinese learners?
4. How can this case study provide generalizations that would be of interest to people working with other endangered languages and the language documentation field in general?

**2. METHODOLOGY.** This section begins with a description of the participants in the Yami language courses, followed by the qualitative and quantitative methods used to analyze the data.

**2.1 PARTICIPANTS IN THE YAMI LANGUAGE COURSES.** The authors of this paper all participated in the process of teaching and learning Yami in various stages of the Yami language courses from 2005-2007. In Spring 2005, a Yami native instructor and a linguist co-taught Yami lessons to five graduate students in linguistics at Providence University in central Taiwan one hour per week as part of the two-credit Austronesian Linguistics Seminar, a graduate elective course, with the goal of introducing the structure of Austronesian languages by providing an authentic example of a language spoken in Taiwan. The same course was repeated in the spring 2006 with six students enrolled. Since half of the learners had a background in ELT, this seminar expanded its coverage from linguistic structures to language use and language teaching and increased the number of hours of Yami instruction from one to two hours per week. All the participants in this course received three graduate credits toward graduation. They were also invited to participate in the research by assessing the Yami e-Learning materials being developed at the same time.

To speed up the process of evaluating all the e-Learning materials,<sup>4</sup> an intensive non-credit language course was offered to four students who had acquired novice-high level of proficiency in Yami<sup>5</sup> to reach the intermediate level, while participating in the project as graduate assistants. The course was held at the National Museum of Natural Science, for a total of 30 hours during eight weekends in the summer of 2006.

After the Yami e-Learning program was finally completed at the end of 2006, two new graduate students signed up for the same Austronesian Linguistics Seminar with two of the three hours devoted to Yami language study in Spring 2007, while two of the graduate assistants with an intermediate level of Yami proficiency continued to participate in the Yami course as researchers and learners.

**2.2. ANALYSIS OF THE QUALITATIVE DATA.** The data for the qualitative studies were collected from 2005-2007, following an ethnographical approach by interviewing the participants, taking field notes, observing and videotaping classroom interactions, and examining the students' assignments and teachers' feedback. The interviews were conducted

---

<sup>4</sup> These are the web-based interactive language learning materials for Yami, available online (<http://yamiproject.cs.pu.edu.tw/elearn>).

<sup>5</sup> See ACTFL guidelines, <http://www.sil.org/LinguaLinks/languagelearning/OtherResources/ACT-FLProficiencyGuidelines/contents.htm>.



by one of the graduate assistants as a participant observer in this project. Themes on teaching pedagogy generalized from the interviews with the instructors and the learners will be discussed in 3.1.

**2.3. ANALYSIS OF THE QUANTITATIVE DATA.** The quantitative data were recorded from the five guided conversations among the four learners once every two to three weeks from March to mid-June 2007 in the Yami language class, which met for 100 minutes per week at Providence University. Each conversation session lasted for 40-50 minutes. The data were transcribed and analyzed to determine phonological and syntactic acquisition processes and variability. For the investigation of the sequences of development, markedness and interference effects were examined. For interlanguage variability, a sociolinguistic variationist approach was used to compare the performance of two groups of learners (intermediate vs. novice) to determine the factors that account for Yami interlanguage variation.

**2.3.1. STATISTICAL ANALYSIS.** The data were analyzed using the VARBRUL program,<sup>6</sup> a loglinear regression analysis suitable for natural language data that usually involves many interacting factors and unbalanced distribution in each cell.

For the study on phonological acquisition, the dependent variables are the four consonantal phonemes identified as difficult for Chinese learners: /k/, /ŋ/, /s/, /z/. The phonological system of Yami will be discussed in 3.2. The independent variables are the learners' proficiency (intermediate vs. novice) and time (five time intervals).

For the study on word order acquisition, the dependent variable is the alternation between VS and SV, while the independent variables are two proficiency levels, five time intervals, four tasks, and two sentence types, the details of which will be discussed in Section 4.

**3. RESULTS AND DISCUSSION.** We begin this section by addressing the first research question on the process of teaching and learning an endangered language with a focus on the teacher's pedagogy.

**3.1 TEACHER'S PEDAGOGY.** Following Borg's qualitative study (1998) on teachers' pedagogical systems and grammar teaching, we grouped the interview data into six major categories: (1) corrective feedback and error analysis, (2) explanation of grammar rules, (3) reference to students' background, (4) grammatical terminology, (5) practicing grammar, and (6) grammar and communicative ability. Table 1 presents the teachers' pedagogy used during the teaching process. The second column lists classroom episodes representing the identified themes<sup>7</sup>, while the third column lists the interview data of the two teachers' com-

---

<sup>6</sup> Paolillo (2002) has provided a comprehensive discussion of the statistical models and methods for analyzing linguistic variation. Tagliamonte (2006) has provided the most detailed step-by-step procedures for VARBRUL analysis and interpretations.

<sup>7</sup> The theme of explanation of grammar rules is illustrated by Episode 2. The theme of reference to students' background is illustrated by the teachers' comments in the third column.

ments on their beliefs about teaching strategies. In the following paragraphs, we combine the interview data with the students' comments to illustrate the process of teaching and learning.

TABLE 1 Teacher's pedagogical systems used during the teaching process

Teacher's pedagogy	Examples	Teacher's comments
<p>Corrective feedback and error analysis</p> <p>Teachers analyzed the students' grammatical errors and provided corrective feedback.</p>	<p><u>Episode 1</u>            S1: mo magaga, maciavang namen pa jimo an?            (Hey, magaga. Can you give us a ride?)</p> <p>S2: kamo mangay jino?            (Where are you (plural) going?)</p> <p>T1: There is only one person.</p> <p>S2: ka...(hesitated) (ka...)</p> <p>T1: You (singular)</p> <p>S2: mo mangay jino?            (Where are you [singular] going?)</p> <p>T1: ka, ka mangay... (ka. ka mangay...)</p> <p>S2: ka mangay jino? (Where are you going?)</p>	<p>Originally, I wanted to use a communicative language teaching method, but I immediately saw that the students struggled greatly and could not keep up. I found later that even in reading a sentence, the students were unable to process it, that they could not process the sentence because we were going too quickly. You had not reached that level, so you couldn't produce those sentences on your own. However, you could understand them passively. Therefore, it was the proper time to use the grammar translation method, so we began to go back to grammar translation, or we would not have been able to continue this course. (T1, December 11, 2006)</p>

<p>Explanation of grammar rules</p> <p>The teachers raised the students' grammatical awareness.</p>	<p><u>Episode 2</u></p> <p>T1: I just said that pasdepen is the main verb, right? A transitive verb. Why is it a transitive verb? You saw that affix, so you say it is a transitive verb.</p> <p>Ss: (quiet)</p> <p>T1: -en, -en. Ok, what about its agent, what is the agent in this sentence?</p> <p>S3: nio. (You.)</p> <p>T1: Nio, You! The agent is "you". "You" is the agent doing the action.</p> <p>What is "you" doing? You pasdepen, so pasdepen means...?</p> <p>T2: Putting something inside.</p> <p>T1: The Ba-construction is used. Do you see that? It is very transitive, isn't it? That's why she used the ba-construction to put something in it. We can find that "pasdepen" is a transitive verb. Ok, what's the patient of this verb? O tatala, right? O tatala nio is the patient semantically; however, it is the grammatical subject in this sentence, because there is an "o" here.</p>	<p>I think first of all, because of their background in English, they were sensitive to grammar, and they knew that grammar is part of the language. Therefore, in terms of grammar, in my observation, they were willing to learn it, and wanted to learn it because it was necessary. This is my observation. (T1, December 11, 2006)</p> <p>What I want to say is that all of you have the linguistic concepts, the language concepts. Thus, you can absorb and understand grammar quickly. (T2, December 17, 2006)</p>
<p>Reference to students' background</p> <p>Teachers changed their pedagogical approach to GTM, based on their perceptions of students' backgrounds and expectations.</p>		

<p>Grammatical terminology</p> <p>Terminology was used quite freely and students coped well with terminology.</p>	<p><u>Episode 3</u></p> <p>T2: apia o tatala namen ang? (What should we do about our boat?)</p> <p>Ss and T1: apia o tatala namen ang? (What should we do about our boat?)</p> <p>T1: Ok, where is the subject?</p> <p>S3: o tatala. (o tatala namen)</p> <p>T1: namen, o tatala namen means?</p> <p>S3: Our boat.</p> <p>T1: Our boat. Ok, and ang?</p> <p>S4: Particle.</p> <p>T1: Particle. Meaning...?</p> <p>S4: (The student gave an inaccurate answer)</p> <p>T1: I am asking about ang right now, what does ang mean?</p> <p>It is the sentence final particle. What's the possible meaning of this word?</p> <p>(silence)</p> <p>T1: Interrogative particle!</p>	<p>For some indigenous language teachers, they have to learn grammar rules. Although you won't teach students these rules, you have to be equipped with the concept of grammar. For example, ma- would appear in front of all adjectives (stative verbs). For example, very tall and very short, etc. It may spark a students' interest in learning the Yami language. (T2, December 17, 2006)</p>
<p>Practicing grammar</p> <p>Teachers always led the course by practicing activities.</p>	<p><u>Episode 4</u></p> <p>Practicing activity: "mangay tamo do_____."</p> <p>T1 provided photos of small boats (one and two-man boats) and big boats (six-man boat, eight-man boat, and ten-man boat) to illustrate the differences. Students asked what ships other people would like to take.</p> <p>T2: mi tamo do mibekbek an, no mangay tamo do atlo so avat?</p> <p>(Should we take the motor boat or the six-man boat?)</p> <p>S4: mangay tamo do pikavangan. (Let's take the two-man boat.)</p>	<p>"I like Audio-lingual because it has strict drills which are stimulating. I like it very much because I think I like this kind of practice. I think it is necessary to have this kind of practice because we can internalize some grammar rules through them. (T1, December 11, 2006)</p>

Grammar and communicative ability  Teachers believed grammar is the basis for communicative ability.	<u>Episode 5</u> Instruction: Pretend that you were one of the five tourists. Write down a list of things that you are going to complain about to the hotel owner. Example 1. ya masazówsaw no mahep, am ya mánnget no maraw. (It is cold at night and hot in the morning.) 2. ya abo o wanglu. (There is no internet access.)	“I think the grammar translation method is the fastest way to know one language in a short time. It is the fastest, because you don’t need to be actively manipulating the language, but can learn the language passively. (T1, December 11, 2006)
--	---	--

**3.1.1. CORRECTIVE FEEDBACK AND ERROR ANALYSIS.** This course was originally designed for communicative language teaching (CLT); however, the teacher T1<sup>8</sup> became aware of students’ difficulties during oral activities and was also concerned about the short timeframe of the course, so she shifted to grammar translation (GTM) and audio-lingual methods beginning with Lesson Two at the beginning stage of her instruction during the 2005-2006 school year. As Maslova (2000) suggests, teachers may change their pedagogies based on the dilemma they face, the teachers’ personal belief systems as well as teaching context and, to a lesser extent, personal background characteristics. The tensions between GTM and CLT were resolved by an eclectic combination of methods and activities, as described in Fotos (2005: 668).

In a change of her teaching strategy, T1’s focus was drawn to corrective feedback and error analysis. Episode 1 in Table 1 provides an example: as soon as a student made a mistake, T1 jumped in to help correct it.

At first, T1 only offered S2 a cue in Chinese “you (singular),” because she wanted to elicit the rule through an interactive discussion rather than simply supplying the rule herself. In fact, S2 offered the correct nominative form “ka” ‘you (singular)’ in her attempt, but after T1’s prompt, she changed to an incorrect answer. After that, T1 gave S2 the correct answer.

T1 mentioned in (1) she liked to offer “incidental corrections” during oral activities, an idea adopted from Ellis’ “incidental focus on form” (2001) as cited in Williams (2005: 671).

(1) What I like is not a shotgun approach, but incidental corrections. I can generalize some common problems emerging from our activities; after that, we can discuss them all together. This is what I like to do. (T1, December 11, 2006)

This preference is drawn from her own experience in learning English inductively in junior high school, as shown in (2).

(2) When I was a junior high school student, my teacher also liked the inductive method. Some teachers prefer to use the deductive method, which is to offer a rule first, and then allow students to put that rule into practice. For me, I don’t like

<sup>8</sup> Throughout this paper, the two teachers are identified as T1 and T2, and the students are identified by the letter S and a number (e.g. S1, S2, etc.).

people to tell me the rules directly since I can find out the rules by myself. For this reason, I prefer to use the inductive method rather than the deductive method. (T1, December 11, 2006)

Sometimes the teacher provided compressed grammar rules rather than asking students to explore the rules themselves, especially when the rules involved transitive and intransitive affixes. As Bauer and Nation (1993) suggest, the majority of affixes are regular and predictable, and therefore should be taught directly. S4 expressed how she felt about this technique in (3).

(3) Maybe she was afraid that we couldn't understand or couldn't answer it, so she told us the rules before asking us. Is that right? If she threw some grammatical questions at you suddenly, you probably would not be able to answer, right? (S4, December 21, 2006)

Although the students felt embarrassed when they were corrected, they still expected corrective feedback, as illustrated in (4).

(4) I felt that the teacher's correction was ok, although I felt embarrassed at that moment. However, if there was no correction, I would not have known my errors when I read out loud. (S1, December 11, 2006)

**3.1.2 EXPLANATION OF GRAMMAR RULES.** During each class, students repeated after T2, modeling the native speaker's input. This was followed by T1's explanation of grammar. This was a recurrent pattern in their team teaching. According to Brown (1994), a very limited portion of classroom speaking time may legitimately be spent in human "tape recorder" speech. Moreover, he also suggests that drills can offer students an opportunity to listen and orally repeat certain strings of language that may pose some linguistic difficulty, either phonological or grammatical. Particularly, these exercises could help to establish certain psychomotor patterns (to "loosen the tongue") and to associate selected grammatical forms with their appropriate contexts.

In Episode 2, Table 1, T1 posed a grammatical question. After several tries, with T2's help, S3 provided correct short answers on case markers. This interaction style is a typical "monologue in disguise", as described in Peled-Elhanan and Blum-Kulka (2006).

A student commented in (5) that teaching grammar before they even learn to pronounce words properly would discourage them from speaking. However, repeating after the teacher encouraged them to speak.

(5) If you teach grammar first before we know how to read (Yami words), and we seem to know everything when we actually do not, then I would be afraid of speaking. If we begin by repeating after the teacher out aloud, then it is a good teaching method. This way, we would be less hesitant to speak when we answer questions or join in activities later. (S2, December 17, 2006)

**3.1.3 REFERENCE TO STUDENTS' BACKGROUNDS.** T1 believed her focus on grammar was related to the students' backgrounds. When the interviewer asked her about her choice of teaching strategy, she mentioned her perceived needs of the participating learners, as in (6).

(6) Most of the students here wanted to finish their MA theses. Take ER<sup>9</sup> for example. He must clearly understand Yami grammar before he can write about word order, right? In addition, you worked on Yami reduplication and KR worked on Yami intonation. JM worked on E-learning and AN worked on Yami dictionary. All of them need to master Yami structures before they can make significant contributions. Especially people like JM - in order to create e-Learning games, she has to pressure cook, cook quickly; she needs to learn the language in a short time. This is why I applied the techniques that I did. (T1, December 11, 2006)

Many adult learners expect grammar in the L2 curriculum and treat grammar as the central component of language (Hinkel and Fotos 2002). Since this was the instructor's assessment of the students' needs, she adopted what she called "pressure cooker" method by focusing on grammar.

**3.1.4 GRAMMATICAL TERMINOLOGY.** Episode 3 in Table 1 shows that the teacher employed a translation strategy and often used grammatical terminology to explain the usage of affixes and case markers in Yami. S3 reflected in (7) the need for translating the Yami sentences into Chinese, especially when the language is written in the Roman alphabet:

(7) Yami is written in Romanization, so when I read the text, I want to know the meanings of the words. (S3, December 9, 2006)

To teach students grammatical terms, T1 employed the above-mentioned "monologue in disguise" or self-questioning and answering technique. In Episode 2, Table 1, she tried to explain the transitive verbal affix "-en" to students. S1 commented that the use of grammatical terminology helped her remember how to use -en, as in (8).

(8) After I learned the meaning of some affixes, it became easier for me to remember them. For example, it is easier to remember the transitive affix -en. (S1, December 11, 2006)

**3.1.5 PRACTICING GRAMMAR.** Another teaching method the teachers used frequently in their course was language activities. In Episode 4, Table 1, T1 provided photos of small boats (one and two-man boats) and big boats (six-man boat, eight-man boat, and ten-man boat) to illustrate the differences. Students asked what ships other people would like to take. After the students learned the topic "*mangahahap*, go fishing," and finished an oral activity and grammar patterns, T1 provided a sentence pattern "*mangay tamo do \_\_\_\_\_*." (We will go by    ) and asked each student to fill in the blank.

Celce-Murcia and Hilles (1988) mention that classroom activities can enable students to perform spontaneously with the language, as well as experience increased empathy, heightened self-esteem and motivation, and lowered sensitivity to rejection, thus facilitating second-language acquisition. S4 expressed positive feelings toward these activities in (9).

---

<sup>9</sup> All names are pseudonyms.

(9) I think it is effective. As a novice, you will be afraid to make a sentence. This way, we can follow the sentence pattern. It is easier to fill in the blank. (S4, December 21, 2006)

T1 also used rhythm and mnemonic devices to increase students' long-term memory. She believed Yami songs could reinforce one's memory of Yami language, so she taught a few songs. A student reflected on the usefulness of songs for language learning in (10).

(10) You will still remember them in the future, for example, I took Japanese in college, and in class we would sing *zo san, zo san* (the elephant song). I still remember it now, so I still know elephant. Thus, it can give you a strong memory, and I think it can stay in my mind for much longer. (S4, December 21, 2006)

She also taught students some effective tricks to remember Yami words such as matching Yami syllables with Chinese characters (see Table 2).

TABLE 2: Yami-Mandarin Chinese phonetic and phonological similarities words

Yami	Mandarin Chinese	English gloss
rangirang	rang-i-rang	rainbow
kasingasingat na	ga-xing-a-xing-a-te-na	It's too expensive

**3.1.6 GRAMMAR AND COMMUNICATIVE ABILITY.** Although T1 said she changed her pedagogical approach to a combination of grammar translation and audio-lingual methods, during classroom observations, many communicative activities were still employed, such as role-plays.

In Episode 5, Table 1, T1 asked the students to pretend that they were one of five tourists and asked them to write down a list of things to complain about to the hotel owners. T1 believed that grammar translation is a short cut to grasp the knowledge of a language. Afterwards, the students can spend the rest of their lives developing communicative ability and improving their accent, as illustrated in (11).

(11) I think that the grammar translation method is the fastest way to know a language in a short time. It is the fastest, because you don't need to be actively manipulating the language, you don't have to speak the language, but can learn the language passively. If you know the grammar rules of that language, you can learn the language passively. (T1, December 11, 2006)

This teacher's practice corresponds with the finding of a study by Lee and Wang (2002) that grammar instruction can make a difference for L2 acquisition and speed up the learning process for adult learners, although communicative approaches have been greatly promoted in language teaching.

**3.1.7 INDIGENOUS LANGUAGE TEACHING IS NO DIFFERENT FROM FOREIGN LANGUAGE TEACHING.** The instructor's teaching practices and the learner's expectations of learning a Taiwanese indigenous language is similar to teaching and learning a foreign language in Taiwan. This may not be an overstating of a minor result. Perhaps



from a psycholinguistic perspective, the process of second language acquisition would be the same regardless of the social status of the language. However, from a sociolinguistic perspective, this finding has many implications and might eventually have an impact on language revitalization.

It might sound ridiculous and politically incorrect to treat indigenous languages as “foreign” in their own land.<sup>10</sup> However, the strategy of teaching an endangered indigenous language to the majority of the society and to position it as a “foreign” language, which carries its own prestige and desirability, is designed to increase the visibility and status of the endangered language.

The language learning effort made by the Chinese students, in fact, has made an impression on the Yami local activists whose children do not speak Yami. In addition, the learners also chat online in Yami, arousing curiosity from their peers to encourage them to sign up for this language course. Therefore, these endeavors to promote Yami by teaching it as a “foreign” language in a graduate program will undoubtedly raise the prestige of the language.

**3.2. PHONOLOGICAL ACQUISITION.** In this and the following sections, we focus on the developmental sequences and interlanguage variability in phonology and syntax.

There are twenty consonants, four vowels, and four diphthongs in Yami (Rau and Dong 2006). Among all the segments, four consonants, /k, s, ŋ, z/, were identified as difficult for Chinese learners. Table 3 lists Yami consonants, where all the symbols represent standard Yami orthography.

TABLE 3: Classification of Yami consonants

		Place of articulation						
		Labial	Alveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Manner	Stop	p, b	t	d		k, g		ʔ
	Fricative	v		s			h	
	Nasal	m	n			ŋ		
	Liquid		l	r				
	Affricate				č, ǰ			
	Trill		z					
	Glide	w			y			

The four phonemes are explained as follows.

/k/: a voiceless stop.

/ŋ/: a velar nasal.

<sup>10</sup> This analogy actually struck an ethnic Amis scholar in Taiwan as unbelievable (p.c., Akio Huang), when “indigenous” language acquisition was considered “foreign” language acquisition.

/s/: a voiceless retroflex fricative. It is palatalized before the front vowel /i/, e.g., *siko* ‘elbow’, *sinsi* ‘teacher, loan word’. The /s/ occurring before any vowel other than a high front vowel /i/ in a loan word is pronounced as a voiceless alveolar fricative [s], e.g., *saki* ‘liquor’.

/z/: an alveolar trill.

Some of the most commonly found variants for the four phonemes are listed in Table 4, with the first row representing the target form.

TABLE 4: Variants for the four phonemes

/k/ → [k]	/ŋ/ → [ŋ]	/s/ → retroflex /s/	/z/ → trill
/k/ → [k <sup>h</sup> ]	/ŋ/ → [n]	/s/ → [s]	/z/ → [l]
	/ŋ/ → [g]		/z/ → [z]
	/ŋ/ → [ʔ]		

The percentages of correct production of the target forms are shown in Table 5.

TABLE 5: Description of the realization of the four sounds

Phoneme		Input Probability Production/ Total tokens	%
/k/		<b>0.98</b>	
	[k]	830/882	94.1
	[k <sup>h</sup> ]	52/882	5.9
/ŋ/		<b>0.88</b>	
	[ŋ]	397/466	85.2
	[n]	7/466	1.5
	[g]	49/466	10.5
	[ʔ]	13/466	2.8
/s/		<b>0.59</b>	
	[retroflex s]	122/207	<b>58.9</b>
	[s]	85/207	41.1
/z/		<b>0.04</b>	
	[trill]	13/282	<b>4.6</b>
	[l]	257/282	91.1
	[z]	12/282	4.3

Among the four problematic phonemes, the trill /z/ is the most challenging (accuracy rate = 4.6%), followed by the retroflex /s/ (accuracy rate = 58.9%). The velar nasal in non-coda position has a wide variety of variants, whereas the problem of the velar stop /k/ lies mostly in aspiration, which might be interference from the English alphabet k.

Among the four phonemes, /z/ is the most challenging sound to acquire. The corrected mean of accurate production or the input probability of /z/ from the VARBRUL analysis is only 0.04, as shown in Table 5. The corrected means of accurate production for /s/, /ŋ/, and /k/, on the other hand, are 0.59, 0.88 and 0.98, respectively.

The alveolar trill /z/ is extremely marked because it is infrequent in languages of the world; besides, neither the learners' L1 (Mandarin Chinese) nor their major foreign L2 (English) have the trill sound in their inventories. Thus according to Eckman's (1996) Markedness Differential Hypothesis (MDH), trill is a difficult sound for the learners to pronounce. The two variants for trill are lateral [l] and retroflex [r], such as *pivazayan* [pivaɻayan] "work, transitive verb" and *mivazay* [mivaɻay] "work, intransitive verb". Accurate production of the target trill did not occur at all during the first year of instruction and thus will not be further discussed in this study.

In our following analysis of sequence of development and variability, we only concentrate on the two variables with the most variants, i.e., /ŋ/ and /s/. A log-linear regression analysis, VARBRUL, was used to test how proficiency and time account for sequence of development and variability.

**3.2.1 INVESTIGATION OF /ŋ/.** The most correct pronunciations of [ŋ] occurred in coda position, for example, *ikong* 'what,' *Maoyong* 'person's name,' *manazang* 'buy.' The learners tended to resyllabify the syllable structure with /ŋ/ by changing it from the onset to the coda<sup>11</sup> and inserting a [g] to create a new onset. For instance, they would pronounce *manga-mi-zing* 'obedient' as *mang-ga-mi-zing*; *ka-te-ngan* 'know' as *ka-teng-gan*.

Four separate VARBRUL runs were conducted on the four variants of (ŋ) with a total number of tokens of 466 and found that proficiency determined accurate production of the target form of (ŋ). As shown in Table 6, students of intermediate proficiency used the target form more frequently (.65), whereas those of low proficiency tended to use the three variants (.78 for [n], .64 for [g] and .78 for glottal stop). Overall, the most frequent variant for the target form (ŋ) is [g] with a total frequency of 10.5% as opposed to a little under 3% for either of the other two variants [n] and glottal stop, as shown in Table 6.

TABLE 6: Proficiency on accurate production of (ŋ)

Factor Group	VARBRUL Weight (%)			
	[ŋ]	[n]	[g]	[ʔ]
<b>Proficiency</b>				
Novice	.32 (76)	<b>.78</b> (3.4)	<b>.64</b> (15.9)	<b>.78</b> (5.1)
Intermediate	<b>.65</b> (92)	.27 (0.4)	.39 (6.2)	.22 (0.4)
Range	34	51	25	56
Total Percentage %	85.2 % (397/466)	<b>1.7 %</b> (8/466)	<b>10.5 %</b> (49/466)	<b>2.8 %</b> (13/466)

<sup>11</sup> The only two consonants permissible in coda position in Mandarin Chinese is /n/ and /ŋ/.

As shown in Table 7, time was also a significant factor group accounting for accurate production of (ŋ). The ratio of correct pronunciation increased gradually; the third and fifth times were better than the first, second, and the fourth times, as the VARBRUL weights above .50 (.60 and .57 respectively) indicate promotion of accurate production. The fact that both proficiency and time determine accurate production of (ŋ) indicates that the development of (ŋ) follows a linear pattern.

TABLE 7: Time on accurate production of (ŋ)

Factor Group	VARBRUL Weight (Pi)	Correct tokens / Total tokens	%
<b>Time</b>			
First	0.26	36/51	70.6%
Second	0.44	80/97	82.5%
Third	<b>0.60</b>	108/120	90%
Fourth	0.48	65/76	85.5%
Fifth	<b>0.57</b>	108/122	88.5%
<b>Range</b>	<b>34</b>		
<b>Total</b>		397/466	85.2%

**3.2.2 INVESTIGATION OF /s/.** The acquisition of (s), on the other hand, demonstrates a different trajectory. Contrary to the linear pattern for (ŋ), there seems to be a curvilinear pattern for (s). Two separate VARBRUL analyses were conducted on the 207 tokens with /s/. All loan words and palatalized variants were excluded from coding, thus the target form is the retroflex /s/. As shown in Table 8. The only significant factor group is time. Beginning with the third time, the learners showed accurate production of the target form with the weights of .55, .66, and .61 respectively. However, the intermediate learners did not favor the target form and the difference between them and the novice learners was not significant<sup>12</sup>.

<sup>12</sup> By convention, the VARBRUL weights for non-significant factors are listed in square bracket.

TABLE 8: Significant factors accounting for accurate production of (s)

Factor Group	VARBRUL Weight (Pi)	Correct tokens	%
		/	
		Total tokens	
<b>Proficiency</b>			
Novice	NS [0.52]	62/102	60.8%
Intermediate	NS [0.48]	60/105	57.1%
<b>Time</b>			
First	0.35	6/15	40%
Second	0.35	18/45	40%
Third	<b>0.55</b>	9/15	60%
Fourth	<b>0.66</b>	45/65	69.2%
Fifth	<b>0.61</b>	44/67	65.7%
<b>Total</b>		<b>122/207</b>	<b>58.9%</b>

Students of novice level actually produced the target form of retroflex /s/ slightly more frequently than did those of intermediate proficiency, who preferred to use [s]. This is probably because there is a phonological variation and sound change in progress between /s/ and the retroflex /s/ in Mandarin Chinese (Rau and Li 1994, Rau 1996). There is a tendency for Taiwan Mandarin speakers to merge the two phonemes to /s/. For the novice-level students, the retroflex feature was more salient to them because their attention was recently drawn to this in Yami. On the other hand, intermediate learners tended to fall back on their L1 habit when they joined the course again.

Previous studies have shown that phonological variation in L1 might affect interlanguage variation in L2. Chen (2001) provided evidence to show that Chinese learners of English who used a stigmatized variant of (r), such as [l], in their L1 also used the same variant for their English (r), e.g., [lwak] and [low] for *rock and roll*. In addition to the variant of [s] for the retroflex target in our data, we also found the variant of Chinese [l] for Yami (r), e.g., a<sub>r</sub>aw ‘day’ [a<sub>l</sub>aw] by an intermediate learner who used a stigmatized variant of (r) in Mandarin Chinese.

**3.2.3 SUMMARY.** We have learned that the four phonemes which tended to be pronounced inaccurately actually have different degrees of difficulty. Even for the two most difficult phonemes, the path of trajectory is different. These findings in phonological acquisition in L2 will serve as useful guidelines for developing materials for teaching Yami pronunciation and contribute to endangered language revitalization, as most of the Yami youth are learning Yami as a second language.

**3.3 ACQUISITION OF WORD ORDER IN YAMI.** In this section, we present the results of word order acquisition in Yami, using the VARBRUL program to analyze 500 tokens of word order data, gathered from four tasks: (1) choosing the correct Yami word order with Chinese translation, (2) choosing the correct Yami word order without Chinese translation, (3) translating sentences from Chinese into Yami, and (4) sentence arrangement test. The

four learners were asked to complete all four tasks in Mid-May 2007 toward the end of the language course. Before the results of word order acquisition are presented, a word on Yami word order is in order.

The unmarked word order in Yami is verb initial (VS) as in (12); however, a pronominal subject is fronted (SV) to mark the present tense (Rau 2005, Rau and Dong 2006) as in (13).

The VS order is unmarked in narratives but the SV order occurs frequently in conversations, the function of which is to mark present tense.

(12) k-om-anso      wakay      si \_\_\_\_\_ Salang.  
<intransitive>eat      Oblique sweet potato      Nominative personal name  
 V      S

‘Salang wants to eat a sweet potato.’

(13) kamo      m-angay      jino  
you (pl.)      intransitive-go      where  
 S      V

‘Where are you going?’

The results of word order acquisition are presented in Table 9. Word order in Yami for Chinese learners is not as difficult as expected, as the percentage of accurate production is 63% (Input probability= 0.63).

TABLE 9: VARBRUL results of word order acquisition

Factor Group	VARBRUL Weight (Pi)	Correct tokens/ Total tokens	%
<b>Level</b>			
Intermediate	<b>0.68</b>	171/250	68.4%
Novice	<b>0.54</b>	135/250	54%
<b>Task</b>			
Sentence Arrangement	<b>0.35</b>	111/236	47%
Sentence choice (with Chinese translation)	<b>0.68</b>	90/116	77.6%
Translation from Chinese to Yami	<b>0.50</b>	20/32	62.5%
Sentence choice (without Chinese translation)	<b>0.63</b>	85/116	73.3%
<b>Total</b>		306/500	61.2%
<b>Input probability = 0.63</b> Total Chi-square = 6.4959			
Chi-square/cell = 0.8120 Log likelihood = -307.427			

Although intermediate students performed a little better than the novice group, both groups tended to provide the correct word order (VARBRUL weights are 0.68 vs. 0.54).

Among the four tasks in Table 9, except for translation from Chinese to Yami, which had no effect on the accurate production of word order (0.50), the other three all affected production of word choice one way or the other. Sentence choice with or without Chinese translation promoted accurate choice of word order (0.68 and 0.63). The sentence arrangement test, on the other hand, tended to inhibit it (0.35).

The questions in sentence arrangement test were further analyzed by running Item-Analysis, following Rau (1999), and only 11 question items with a facility index above .50 and a discrimination index of 1 were chosen to form the instrument for further analysis.

The internal consistency reliability of the question items was also analyzed. The overall Cronbach's alpha is 0.98, which is much higher than the acceptable rule of thumb of .70, set by Nunnally (1978). In other words, the revised sentence arrangement test has a high reliability.

Table 10 shows the implicational scale of the learners' accuracy in the eleven question items. The two intermediate students answered all eleven questions somewhat correctly,<sup>13</sup> while the two novice learners could not arrange any of the eleven sentences into the correct order. Moreover, according to the implicational scale in Table 10, we can predict that Q5, Q16, Q27, Q39 are easier than Q14, Q19, Q26, Q28, Q47, Q51, Q52. (See Appendix: Sentence arrangement test)

TABLE 10: The implicational scale of the students' Yami language acquisition with the eleven questions

Question Subject	Q5	Q16	Q27	Q39	Q14	Q19	Q26	Q28	Q47	Q51	Q52
<i>DV</i>	1	1	1	1	1	1	1	1	1	1	1
<i>KR</i>	1	1	1	1	0	0	0	0	0	0	0
<i>CR</i>	0	0	0	0	0	0	0	0	0	0	0
<i>TR</i>	0	0	0	0	0	0	0	0	0	0	0

All question items came from Yami lessons at the beginning level. Since the first four questions were selected from the early lessons, they tended to be familiar to and memorized by the learners. Once the word order departed from the SV, it posed a challenge to our beginning learners.

To test whether novice learners tend to use the SV order whereas intermediate students are more willing to try the VS order, we coded the word order of all the utterances in the five guided conversations as the dependent variable with proficiency level and sentence type as the independent variables. The results are discussed in the following section.

<sup>13</sup> DV created the sentence arrangement test and thus practice effect might have partially explained his perfect score on word order.

**3.3.1 VS / SV VARIATION.** A VARBRUL analysis was conducted on the 494 tokens from the guided conversations. The results indicate that the students' proficiency level and sentence type (interrogative vs. declarative) were selected to be significant factors that can account for the choice of word order between VS and SV.

TABLE 11: VARBRUL results of Yami VS word order production

Factor Group	VARBRUL Weight (Pi)	Tokens with VS	
		/	%
		Total tokens	
<b><u>Level</u></b>			
Intermediate	<b>0.55</b>	175/271	65%
Novice	<b>0.45</b>	118/223	53%
<b><u>Sentence type</u></b>			
Interrogative	<b>0.74</b>	168/204	82%
Declarative	<b>0.32</b>	125/290	43%
<b>Total</b>		293/494	60%
Input probability = 0.617		Total Chi-square = 0.2650	
Chi-square/cell = 0.0662		Log likelihood = -291.339	

As shown in Table 11, the intermediate level learners favored the usage of VS order (0.55), while the low level learners preferred the SV order (0.45). In addition, the occurrence of interrogative sentences strongly promoted the VS order (0.74), whereas declarative sentences favored the SV order (0.32). The following adjacency pair of question-answer in (14) at Time 2 illustrates the use of VS order in an intermediate student's question on whether the interlocutor (her Yami name is Magaga) would like to eat a sweet potato, but a novice student responded inaccurately with the SV order. Her answer is "I dare to eat the sweet potato," rather than "I would like to eat a sweet potato."

(14a) DV: mo Magaga, koman ka so wakay?  
V S

"Would you like to eat a sweet potato, Magaga?"

(14b) TR: nohon, ko makan o wakay.  
S V

"Yes, I dare to eat the sweet potato."

(The correct form should be koman ko so wakay "I would like to eat a sweet potato.")  
V S

The following pair in (15) also illustrates the misuse of SV order by an intermediate student addressing a novice student (his Yami name is Maoyong) in Time 4. Both (14) and (15) demonstrate that the learners actually had difficulty marking the distinction in tense by word order.



(15a) DV: mo Maoyong, ko mitkeh do jia an.

S V

(The correct form should be mitkeh ko)

V S

“I will sleep here, Maoyong.”

(15b) CR: nohon,

“OK.”

In summary, the findings of acquisition of word order by Chinese learners seemed to indicate the SV order is acquired before the VS order. This might be accounted for by L1 transfer as the Chinese word order is SV. In addition, the learners associated word order with sentence type in their interlanguage rather than with tense distinction in the target language. This might be a case of exemplar-based learning (Bybee and Thompson 2000) in the interlanguage system where high type frequency in the input (as reflected in the contents and sequence of the teaching materials) led the beginning learners to hypothesize that declarative sentences were used frequently with SV order whereas interrogative sentences were used frequently with VS order. As with the findings in phonological acquisition, these findings in syntactic acquisition will not only be useful in sharpening linguistic description of Yami but also insightful in developing teaching materials for Yami word order.

**4. APPLICATIONS OF THE YAMI CASE TO OTHER PROGRAMS OF LANGUAGE DOCUMENTATION AND REVITALIZATION.** What can people working with other endangered languages learn from the Yami case? We believe the Yami language program has provided an integrated model for endangered language documentation, revitalization and pedagogical research, which can be implemented in two stages for language planning.

In the first stage of the language planning program, endangered language documentation (Himmelmann 1998, Woodbury 2003) should go hand in hand with development of teaching materials, especially in computer-based instruction (Yang and Rau 2005). The participation of three groups of people is crucial in this process, as described in Rau and Yang (to appear). The community members should be brought into the picture early to facilitate continued partnership of the community members with the research team. The university researchers are primarily content providers and e-Learning developers, with the potential goal of becoming learners. The targeted learners are university students, focusing primarily on graduate students who are interested in languages and cultures, with the potential of extension to the endangered language community members who are two or more generations removed.

In the second stage of the program after the infrastructure has been built, the launch of teaching and learning of the endangered language should go hand in hand with pedagogical research as a strategy for language revitalization. The activities of teaching and learning an endangered language in the university setting will undoubtedly raise the visibility and pres-

tige of the language. Furthermore, the process of pedagogical research will empower the participants of the language program. Finally, the results of the research will feed into the contents of documentation and development of teaching materials to strengthen the cycle.

**5. CONCLUSION.** This study presented the results of a longitudinal study of adult learners' acquisition of an indigenous language in Taiwan. We found that learning an endangered language by Chinese learners as a "foreign" language in Taiwan will undoubtedly raise visibility of the language and promote revitalization. The results on phonological and syntactic acquisition will guide the development of teaching materials to further strengthen teaching and learning of the endangered language.

This study of acquisition of an endangered Austronesian language demonstrates cogently that language is a complex adaptive system (Ellis and Larsen-Freeman 2006). In phonological acquisition, the trill was the most difficult phoneme to learn. It was not usually acquired within the first year, but was usually replaced by lateral [l] or retroflex [r]. We also found systematic variation for the variables (ŋ) and (s), constrained by both markedness and interference. Furthermore, learners tended to interpret Yami orthography based on their knowledge of English and thus pronounced /k/ as aspirated [k<sup>h</sup>] and trill /z/ as [z].

In word order acquisition, learners performed much better than expected, partially because the present tense, coded by the SV word order, is the norm in Yami conversations, and thus we did not test if the learners could distinguish between VS and SV in narratives. However, we still found students inaccurately associated word order with sentence type rather than with tense distinction, probably an artifact of the curriculum. Future studies on word order variability should collect narrative data from advanced learners.

Overall, the research presented in this paper makes a contribution both to second language acquisition research and to language revitalization since it brings together two distinct but crucial areas. The program highlights a much-needed area of language research that integrates endangered language revitalization with pedagogical research – in this new area, researchers from both sides need a step-by-step guide to resolving the issues they will face.

## APPENDIX

Sentence arrangement test

Q5: malavayo sira pa ina ya na?

① ② ③ ④ ⑤ ⑥

Are her parents still young?

Correct answer: ya pa malavayo sira ina na?

Q16: mangay kamo jino?

① ② ③

Where are you going?

Correct answer: kamo mangay jino?

Q27: romiag an kamo pa?

① ② ③ ④

Would you still like to have breakfast?

Correct answer: romiag kamo pa an?

Q39: ya aro o ya mibozo.

① ② ③ ④ ⑤

Many people are playing basketball.

Correct answer: ya aro o ya mibozo.

Q14: vazay ikong o mo?

① ② ③ ④

What is your job?

Correct answer: ikong o vazay mo?

Q19: pa ji do kamo takey nimangay?

① ② ③ ④ ⑤ ⑥

Didn't you go to the mountains?

Correct answer: kamo pa ji nimangay do takey?

Q26: kamo na niromiag?

① ② ③

Did you have breakfast yet?

Correct answer: kamo na niromiag?

Q28: ciaha, rana namen ta mabsoy.

① ② ③ ④

Don't worry. We are full (satiated).

Correct answer: ciaha, ta namen mabsoy rana.

Q47: citoai na am, ko pisikingen inio.

① ② ③

I am going to give you a test in a moment.

Correct answer: citoai na am, pisikingen ko inio.

Q51: ipivatvatek pa ko jimo so mamood an?

①            ② ③ ④   ⑤   ⑥       ⑦

May I borrow a pen from you?

Correct answer: mamood ko pa jimo so ipivatvatek an?

Q52: nivatvatkan malas ya ya ko.

①            ②   ③ ④ ⑤

I wrote this wrong.

Correct answer: ya malas ya ko nivatvatkan

## REFERENCES

- BAUER, LAURIE and PAUL NATION. 1993. Word families. *International Journal of Lexicography*, 6, 253-279.
- BROWN, H. DOUGLAS. 1994. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Prentice Hall Regents.
- BORG, SIMON. 1998. Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly*, 32(1): 9-38.
- BYBEE, JOAN and SANDRA THOMPSON. 2000. Three frequency effects in syntax. *Berkeley Linguistic Society* 23: 65-85.
- CELCE-MURCIA, MARIANNE and SHARON HILLES. 1988. *Techniques and Resources in Teaching Grammar*. New York: Oxford University Press.
- CHEN, SHU-CHUAN. 2001. *Interlanguage variation in the production of English (r) by Taiwanese learners*. MA thesis, Providence University, Taiwan.
- CHEN, HUI-PING. 1998. *A sociolinguistic study of second language proficiency, language use, and language attitude among the Yami in Lanyu*. MA thesis, Providence University, Taiwan.
- ECKMAN, FRED R. 1996. A functional-typological approach to second language acquisition theory. In *Handbook of Second Language Acquisition*, ed. by William C. Ritchie and Tej K. Bhatia, 195-211. New York: Academic Press.
- ELLIS, ROD. 2001. Investigating Form-focused Instruction. *Language Learning*, 51(Suppl.1), 1-46.
- ELLIS, NICK C. and DIANE LARSEN-FREEMAN. 2006. Language emergence: Implications for applied linguistics-Introduction to the special issue. *Applied Linguistics* 24(7): 558-589.
- FISHMAN, JOSHUA. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon (UK): Multilingual Matters.
- FOTOS, SANDRA. 2005. Traditional and grammar translation methods for second language teaching. In *Handbook of research in second language teaching and learning*, ed. by Eli Hinkel, 653-670. Mahwah, NJ.: Lawrence Erlbaum Associates.
- GORDON, RAYMOND G. JR. 2005. *Ethnologue: Languages of the world*. Fifteen Edition. SIL International.
- HENRY, GEORGE and ROBERT ZERWEKH. 2002. SEAsite: Web-based interactive learning resources for Southeast Asian languages and cultures. *CALICO Journal* 19(3): 499-512.
- HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-195.
- HINKEL, ELI and SANDRA FOTOS. eds. 2002. *New perspectives on grammar teaching in second language classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.
- HORNBERGER, NANCY H. 2003. *Continua of biliteracy: An ecological framework for educational policy, research, and practice in multilingual settings*. Clevedon: Multilingual Matters.
- HOVEN, DEBRA. 2003. Strategic uses of CALL: What learners use and how they react. *Australian Review of Applied Linguistics* 26(17): 125-148.
- HUANG, LILLIAN M. 2007. Strategies in revitalizing indigenous languages in Taiwan. Invited paper at the *International Conference on Austronesian Endangered Language Documentation*, June 5-6, Providence University, Taiwan.
- KRAUSS, MICHAEL. 1992. The world's languages in crisis. *Language* 68: 4-10.

- LEE, CHU-TAI H. and CHAOCHANG WANG. 2002. The Effects of teaching a difficult grammatical feature of English through grammar instruction and a communicative approach, *Concentric*, 28 (2): 175-192.
- LIN, YI-HUI. 2007. *A sociolinguistic study on Yami language vitality and maintenance*. MA thesis, Providence University, Taiwan.
- MACLAGAN, MARGARET, JEANETTE KING, RAY HARLOW, CATHERINE WILSON, and PETER KEEGAN. 2006. Language revitalization and Maori – Young L1 and L2 speakers. Paper presented at the NWAV 35, November 9-12, in The Ohio State University.
- MASLOVATY, NAVA. 2000. Teachers' choice of teaching strategies for dealing with socio-moral dilemmas in the elementary school. *Journal of Moral Education* 29: 429-444.
- McFARLAND, CURTIS D. 2006. *A CAI program for teaching Filipino*. Paper presented at the 10<sup>th</sup> International Conference on Austronesian Linguistics, January 17-20, in Palawan, the Philippines.
- NUNNALLY, JUM C. 1978. *Psychometric Theory*. McGraw Hill, New York.
- PAOLILLO, JOHN C. 2002. *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: CSLI publications.
- PELED-ELHANAN, NURIT and SHOSHANA BLUM-KULKA. 2006. Dialogue in the Israeli classroom: Types of teacher-student talk. *Language and Education* 20(2): 110-127.
- RAU, D. VICTORIA. 1995. *Yami language vitality* (in Chinese), Paper presented at the Conference on Language Use and Ethnic Identity, June 16, Institute of Ethnology, Academia Sinica, Taipei.
- RAU, D. VICTORIA. 1996. Applications of variation studies to Chinese language teaching (in Chinese), *Proceedings of Journées d'étude internationales sur l'enseignement du chinois moderne*, 117-126. Association Française des Professeurs de Chinois, Paris: Librairie le Phoenix.
- RAU, D. VICTORIA. 1999. Validity and reliability of grammar proficiency test and evaluation of the effectiveness of freshman English composition class, *Journal of Humanities* 11: 155-184, Providence University.
- RAU, D. VICTORIA. 2005. [http://www1.pu.edu.tw/~dhraru/writing/pdf/Iconicity\\_and\\_Tense,\\_Aspect,\\_Mood\\_Morphology\\_in\\_Yami\\_2005.pdf](http://www1.pu.edu.tw/~dhraru/writing/pdf/Iconicity_and_Tense,_Aspect,_Mood_Morphology_in_Yami_2005.pdf) Iconicity, tense, aspect, and mood morphology in Yami, *Concentric: Studies in Linguistics* 31(1): 65-94.
- RAU, D. VICTORIA. and MING-CHIEH LI. 1994. Phonological variation of (ts), (tsh), and (s) in Mandarin Chinese, Paper presented at the 23<sup>rd</sup> Annual Conference on New Ways of Analyzing Variation (NWAV), October 20-23, Stanford University.
- RAU, D. VICTORIA and MENG-CHIEN YANG. 2007. Teaching and Learning an Austronesian Language in Taiwan. In D. Victoria Rau and Margaret Florey (eds.). *Documenting and revitalizing Austronesian languages* (Special volume of) *Language Documentation and Conservation*. University of Hawai'i Press.
- RAU, D. VICTORIA and MENG-CHIEN YANG. (to appear). Digital transmission of language and culture. In *Language endangerment and maintenance in the Austronesian region*. ed. by Margaret Florey. Oxford: Oxford University Press.
- RAU, D. VICTORIA. and MAA-NEU DONG. 2006. *Yami Texts with Reference Grammar and Vocabulary*. *Language and Linguistics*, Institute of Linguistics, Academia Sinica, Taipei, Monograph A-10.
- RAU, D. VICTORIA, MENG-CHIEN YANG, and MAA-NEU DONG. 2007. Endangered language documentation and transmission. *Journal of National Council of Less Commonly Taught Languages (NCOLCTL)*, University of Wisconsin at Madison, 53-76.

- SLAUGHTER, HELEN B. 1997. Indigenous language immersion in Hawaii: A case study of Kula Kaiapuni Hawaii, an effort to save the indigenous language of Hawaii. In *Immersion education: International perspectives*, eds. by Robert Keith Johnson and Merrill Swain, 105-130. New York: Cambridge University Press.
- TAGLIAMONTE, SALI A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- WILLIAMS, JESSICA. 2005. Form-focused instruction. In *Handbook of research in second language teaching and learning*, ed. by Eli Hinkel, 671-691. Hillsdale, NJ: Lawrence Erlbaum.
- WOODBURY, ANTHONY. 2003. Defining documentary linguistics. In *Language documentation and description Volume 1*, ed. by Peter K. Austin, London: SOAS.
- YANG, CHIA-CHI 2007. *Understanding users' perceptions: Using needs analysis to developing Yami language website*. MA thesis, Providence University, Taiwan.
- YANG, MENG-CHIEN and D. VICTORIA RAU 2005. An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan. Paper presented at the ALR-2005, October 14, Jeju Island, Korea.

D. Victoria Rau  
dhrau@pu.edu.tw

Hui-Huan Ann Chang  
ann.ronald@msa.hinet.net

Yin-Sheng Tai  
wratp2@msn.com

Zhen-Yi Yang  
a8505243@yahoo.com.tw

Yi-Hui Lin  
jazmin.lin1981@gmail.com

Chia-Chi Yang  
nadia691116@hotmail.com

Maa-Neu Dong  
tong@mail.nmns.edu.tw

# 10

## WeSay: A Tool for Engaging Native Speakers in Dictionary Building

Eric Albright and John Hatton

*Payap University  
SIL International*

This paper introduces WeSay, an open source software application designed to involve language community members in the description and documentation of their language. Intended for rugged, low-power hardware, WeSay's simplified user interface removes many barriers that typically prevent the direct involvement of community members. In this paper, we describe the dictionary-building features of WeSay that allow a linguist to tailor a sequence of language documentation tasks to engage community members. These tasks reduce a production step to its simplest form, enabling focused training and division of labor. Word gathering tasks use semantic domains, word lists, or patterns of likely words to build up the dictionary. Successive tasks add specific content, such as glosses and example sentences, to the entries. In addition, the program can prepare simple paper publications designed to promote community support for the effort and can transfer the raw data to the linguist for further processing with tools that are more powerful.

**1. INTRODUCTION**<sup>1</sup>. Like many of the tasks of language documentation and description, creating a dictionary is arduous work. Yet the greatest resource, the members of under-resourced language communities themselves, often remain untapped. Our experience has shown that, when given appropriate tools, gifted community members are often eager to be involved and to contribute.

One might ask why it is desirable to involve community members in language work. On top of the evident rightness of helping interested community members play whatever role they can, we are struck with the fact that in some language communities, members are already attempting it. They may be diving in with inappropriate tools leading to a frustrating inability to finish, but they are doing what they can.

Recognizing an opportunity to expand the quality and quantity of language documentation collected by native speakers, SIL International and the Linguistics Institute of Payap University in Chiang Mai, Thailand are developing a computer application called WeSay<sup>2</sup>. WeSay is a simple dictionary-building tool for use by native-speakers of under-resourced languages. We have divided this paper into four parts. First, we give our perception of the current state of dictionary building by native-speakers. Next, we describe the personas that WeSay aims to serve. A list of the design goals of WeSay follows, and we finish with a walkthrough of WeSay and its configuration tool.

<sup>1</sup> The authors would like to thank the anonymous reviewers for their comments that helped to clarify the paper.

<sup>2</sup> WeSay is open source under MIT license. Downloads, videos, and source code are available from <http://www.wesay.org>



**2. WHY ANOTHER DICTIONARY PROGRAM?** As occasional consultants to those building dictionaries in Southeast Asia and Papua New Guinea, we have seen projects use a number of dictionary building techniques. In this section, we will look briefly at the advantages and difficulties of five of these.

As an easy way to get started, paper is very attractive. Dictionary Development Process (Moe 2006) workshops often produce tall stacks of paper. These may have a semantic domain written at the top, with column for vernacular and national language glosses. Difficulties arise when people want to actually do anything with all that gathered data. If they want to add more fields, correct the spelling, or make a publication, at some point, someone has to type it all in.

Other projects we have seen have made the very expensive mistake of trying to build a large dictionary by typing entries directly into a word processor in an approximation of published form. That is, they try to type in each entry using character formatting, as it looks in printed dictionaries. While this is easy to start, it is quite hard to finish. Different typists produce different formats, producing brittle documents that become unmanageable as soon as the font size or paper size need to be changed. When the programs they used (e.g. Page-Maker) are discontinued and don't run on current computers, all of their hard work may be inaccessible. Recently, colleagues of ours have been asked to write a program to scrape the content out of a 20,000-word document featuring all these problems.

Other projects have used a humble spreadsheet. Each field is given a column. All the words that begin with a particular letter are grouped onto a single sheet within the document. This is a relatively reasonable approach for simple structures. This approach becomes unwieldy when the number of fields grows or more complex structures are introduced, such as multiple senses or example sentences. Complex structures are not part of the goal for many projects. The primary problem comes when the team desires to produce a publication. To the extent that the typists have been consistent in their use of columns, converting the system to one that can be published is not a big obstacle.

Systems based on Standard Format Markers (e.g. Shoebox) have long been a mainstay of dictionary development. While many consider this approach to be without significant fault, as consultants, we often tend to the weary and wounded practitioners of this approach, linguist and non-linguist alike. Some of the common problems we see include inconsistent use of cryptic backslash codes, multiple overlapping versions of dictionary files, lost settings files, a mix of Unicode and non-Unicode encodings within the same files, training requirements that can take weeks, and data stranded in one aging tool by a lack of consistently applied markup standards.

Finally, there are systems designed specifically for entering lexical data. These take responsibility for ensuring the data is well formed by presenting the user with forms to be filled out. Some of these are TshwaneLex (Joffe 2004) & SIL's FieldWorks Language Explorer (Butler and van Volkinburg 2007). These are quite powerful, but they are aimed at a very different sort of user than those we want to empower.

**3. PRIMARY PERSONAS OF WESAY.** In designing WeSay, we have found it useful to use what are called primary personas in the practice of User Interaction Design (Cooper 1999:137). A persona is an archetype who represents the actual users. Actual users will

have varying characteristics, but by meeting the needs of these imaginary persons, the software should satisfy a wide range of people. WeSay has identified two such primary personas.

**3.1 ADVISOR.** Our advisor persona is comfortable with everyday computer tasks, and can get occasional help from more experienced friends and consultants. She may be primarily working in literacy, translation, or lexicography and may be interested in producing language-learning dictionaries, or linguistic-oriented dictionaries. While she may not be a linguist, per se, she has enough training to guide the process of basic dictionary making.

**3.2 USER.** Our user persona is a native speaker of the language being documented. He lives in a village, but can travel to town occasionally and can meet with his advisor several times a year. While his educational opportunities have been limited, he is bright and the limited opportunities of rural life have not challenged his potential. He has no prior computer experience. What education he has had was in a second language, in an educational system that emphasizes rote learning rather than problem solving. While literate in the national language, he has rarely read or written anything in his own language, and so is unsure of spelling. He is motivated by a desire to help his people, and helped by regular encouragement, especially if his local community appreciates his efforts. The community can be a big help, but it can also act as an insurmountable obstacle to his work.

**3.3 OTHER PERSONAS AND SITUATIONS.** In taking this approach to software design and choosing these personas, we must consciously choose not to serve others. Though WeSay seeks to be of use to some groups, we make no claims about its usefulness to many or even most communities. For example, there are groups for which collaborative, web-based tools would be needed to overcome hurdles of distance and the desire for multiple people to contribute simultaneously. There are others for which there is a significant potential for dispute over the contents of the dictionary. This version of WeSay lacks explicit support for tracking who entered (or said) what, and what area they were from (though one can make custom fields to hold these kinds of data). If conflict resolution is needed, it will have to happen outside of the tool.

There are situations in which the would-be dictionary builder is not actually fluent in the language. WeSay could be of use to this person, but it is not designed with their needs in mind. Finally, there are important questions related to language revitalization efforts. How would a dictionary be used, if at all (Corris et al. 2002)? How might a tool be designed to encourage the widest possible community participation in the process of compiling a dictionary, under these circumstances? At this time, we do not have anything to contribute to these important questions.

**4. GOALS.** In designing WeSay, we had five major goals: (1) it should be simple enough to be used by people with little computer savvy; (2) it should support a task-based approach to dictionary creation; (3) it should keep the training load to a minimum by encouraging just-in-time training; (4) it should help promote community support for the process; and (5) it should run on inexpensive, rugged, low-power hardware. In the following sections, we consider each of these.

**4.1 DESIGNED FOR NATIVE SPEAKERS OF UNDER-RESOURCED LANGUAGES.** WeSay's primary distinction is that it has been specifically designed for relatively unskilled native speakers of under-resourced languages, empowering them to be active contributors and creators of dictionaries. Rather than taking already existing tools and stripping them down, we believe that tools that truly support these speakers must be designed from the ground up with their unique concerns and strengths in mind.

By design, WeSay supports only a subset of the task of dictionary creation. We have been able to keep it simple by omitting features, which, though they would be requirements for field linguists, are not requirements for native speakers. We view WeSay as a type of satellite application that can gather data. The data can then be processed by the more powerful tools if warranted by the goals of the project<sup>3</sup>. We would encourage linguists to continue to use tools such as Toolbox and Fieldworks Language Explorer for those aspects of lexicography that WeSay does not support.

Because many of these potential users of WeSay know little about computers, WeSay avoids the complexity associated with other software tools that have been designed primarily for expatriate linguists. For example, even computer savvy linguists often get confused and make significant errors when dealing with the computer's file system and file management tasks (such as finding, opening, saving and backing-up files). Someone may appear to do fine during training; but a month later, work has come to a halt. Users of Toolbox often report that they have lost all their work when, in fact, they just closed a single window and have no idea how to get it to open in their project again.

WeSay handles file management for the user automatically. When the user starts WeSay, he is back where he left off, editing the same file as before. Saving is constant and automatic. Backing up is both invisible and automatic. When the user desires to put all his work on a USB key, email it, or push it up on the web, a single click will do the job.

WeSay also has a very simple user interface. It avoids many common user interface conventions that add complexity, such as menu bars, dialog boxes, and many buttons. As for the data entry itself, the user is presented with forms to fill out, rather than being asked to enter cryptic backslash codes and maintain some abstract entry structure themselves.

To further simplify the user's experience, we locate all configuration controls in a completely separate program: the WeSay Configuration Tool. While still aiming to keep configuration easy, in this tool we put the full set of user interface widgets at the advisor's disposal. The advisor can rapidly switch back and forth between the Configuration Tool and the WeSay experience she is molding for the user.

Finally, we have kept the process and user interaction simple by focusing the user on a single task at a time. Each task has a user interface tailored to the job at hand.

---

<sup>3</sup> WeSay data is stored as a Lexical Interchange Format XML file (Hosken 2006) and can be exported to MDF (Multi-Dictionary Formatter) style Standard Format Markers (Coward and Grimes 1995).

**4.2 SUPPORTS TASK-BASED DICTIONARY CREATION WORKFLOW.** Another major distinction of WeSay is its task-based approach to dictionary creation. By task, we mean not only the user's undertaking, but also a software screen that maximally aids the user's productivity and enjoyment.

Traditionally, the task of lexicography has been tackled by working on single entries. Lexicographers fill out as much information as they can about a single entry, then move on to the next one. As new information is discovered, they may come back to an entry but the intent is to fill out an entry as completely as possible. In contrast, task-based dictionary creation focuses on adding only a single part of an entry at a time, but doing so across all existing entries in the dictionary.

Ron Moe (2001) describes a paradigm shift in which mass production techniques are leveraged by lexicographers. He suggested speeding up the process of dictionary creation with his Dictionary Development Process (Moe 2006) by using semantic domains to elicit words and then to expand the dictionary field by field. Those who have used this process have been excited about it (Shore and van den Berg 2006). WeSay's "Gather by Semantic Domains" task joins SIL's FieldWorks Language Explorer<sup>4</sup> in providing specific tool support for this process.

When we break down the big job of making a dictionary into the smallest set of steps, we are left with a series of tasks. Some of these would be:

thinking up words to be included in the dictionary,

differentiating the senses of the words,

glossing the words into another language,

adding examples to the senses, and

defining the words.

Many advantages come of this task-based approach. In the traditional model, people working on the dictionary had to have the training and entire set of skills necessary to complete any dictionary entry. With a task-based approach, the task can be performed by the person best suited for the job. Some tasks, such as glosses or translations, will require bilingual speakers, but many tasks, such as identifying words and providing example sentences can be performed by monolinguals. Different people may be involved at different times; one person may write examples and later another translates these. Because the bar is lowered, more people can participate.

---

<sup>4</sup> Like WeSay, Fieldworks Language Explorer supports Ron Moe's *Dictionary Development Process* by having a specialized input mode for gathering words by semantic domains. It also has a very powerful *bulk editing* tool (SIL International n.d.) which allows the user to automatically generate some fields based on the content of other fields (such as the tone pattern of the word, CV pattern of the word, or even the grammatical category for languages that have morphology that uniquely identifies the grammatical category).

**4.3 KEEPS THE TRAINING LOAD TO A MINIMUM.** By designing WeSay for native speakers and keeping a task focus, we aim to minimize the training required. For example, by hiding the computer's file system, we render unnecessary this major training area.

What really should go in each field? What is a definition, versus a gloss? The advisor can reduce training in terminology by translating WeSay's user interface into a language users can read, and can pick terms that will communicate well for them. This translation does not require interaction with the WeSay developers; the advisor does it using standard open source localization tools.

WeSay's task focus facilitates just-in-time training. A user can be trained only in the task at hand, rather than the entire dictionary creation process. The trainee is more likely to learn the material well and less likely to be overwhelmed by the concepts. He is less likely to forget what he was taught because he immediately puts it into practice.

**4.4 ENCOURAGES COMMUNITY SUPPORT.** We have designed WeSay with the understanding that the job of language description and documentation can engage not just gifted individuals, but the communities in which they live.

In many communities, someone with outside contacts and a computer will be knocked down, unless the benefits of the activity are seen to accrue to everyone (or at least the right people). With this in mind, WeSay supports making simple printouts of the dictionary and printouts of reports containing charts. If a printer is available, these tangible products can be shared in the community on a regular basis. By giving regular, smart-looking reports to community leaders, the dictionary builder can show respect for authorities, fostering both accountability to them and support from them.

One factor in community participation is simply whether or not the work is done within the community. When software is complex, users can quickly lose confidence, if not data. Thus there is a tendency to try to get the work done with computer help close at hand, in a large city or in a workshop environment. To the extent that a user of WeSay can be successful over sustained periods without such help, he has more freedom to do the work within the community, and over longer periods of time. Since the computer is taken to the community, more members of the community can get involved, whether by sitting with the WeSay user as he goes through tasks, reviewing printouts, or providing lists of words (on paper) grouped by semantic domains.

**4.5 RUNS ON LOW-POWER COMPUTERS.** Many people prepared to work on language development are located in rural communities. Many of these communities have dusty or salt-laden air, which shortens normal computer lifetimes. Many lack mains or even generator-produced power. Fortunately, it seems that these obstacles to community involvement are on the verge of being overcome. While it currently runs on Microsoft Windows machines, WeSay should eventually run on the coming generation of Linux-based laptops that are low-power, low-cost, and rugged. For example, MIT's One Laptop per Child project (OLPC) may well furnish a perfect platform for using WeSay in environments previously off-limits to computers. Costly hardware can still be a problem, even in situations where power and ruggedness are not an issue, so we are putting a lot of effort into ensuring that performance is still acceptable even using old hardware.

**5. A USAGE SCENARIO.** Let us take a look at how a community may develop a dictionary using WeSay. We have separate tools for each of the personas that we introduced earlier: the user, a member of the community who desires to help build a dictionary, and the advisor, the project manager who may be a member of the community or an outsider.

To get started, let us say that we are beginning with a blank slate, with no existing dictionary work to build on. The advisor begins by launching the WeSay Configuration Tool to create a new project.

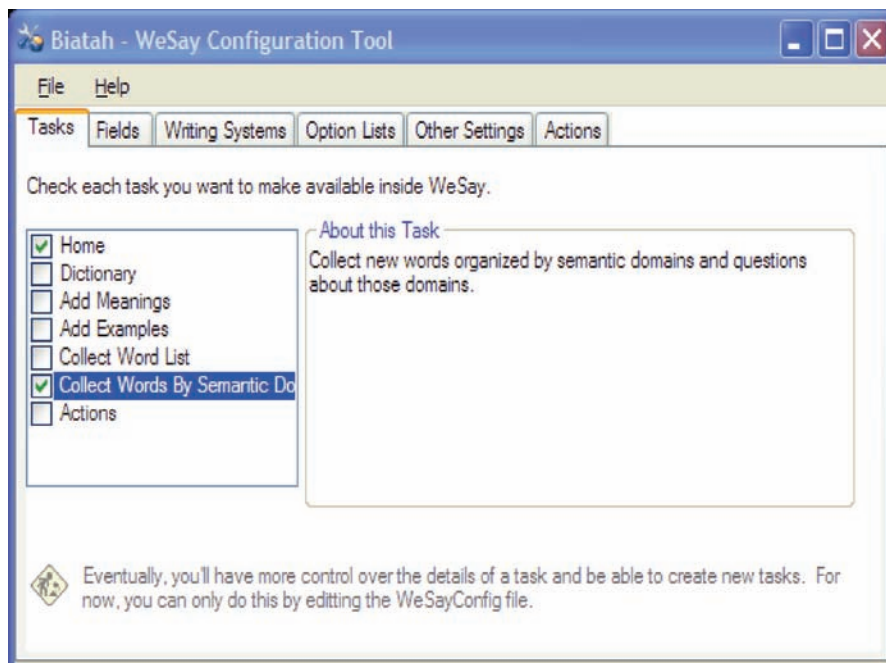


FIGURE 1

She sets up the writing system<sup>5</sup> information, and then decides that the first step will be to collect words using semantic domains. She turns on that task and disables all the others. At this point, she can launch WeSay to see what it will look like for the user (Figure 2).

---

<sup>5</sup> WeSay has support for multiple scripts per language, including complex non-roman scripts. When we say “writing system”, then, we mean a particular way of writing a particular language; such as Simplified Chinese, Traditional Chinese, and Pinyin.

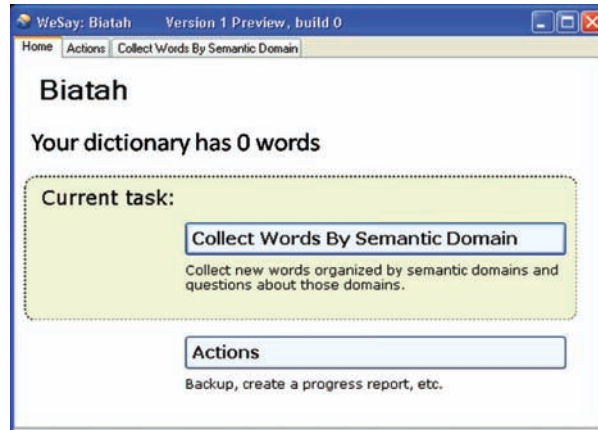


FIGURE 2

Satisfied, she trains the user on collecting words by semantic domains and they practice doing that using WeSay. She provides the user with a couple USB flash drives for backup, shows him how to plug them in, and which button to click on in WeSay to start the backup.

For the semantic domain word collection task, the user, perhaps in cooperation with a group from the community, thinks of words associated with a semantic domain (Figure 3).

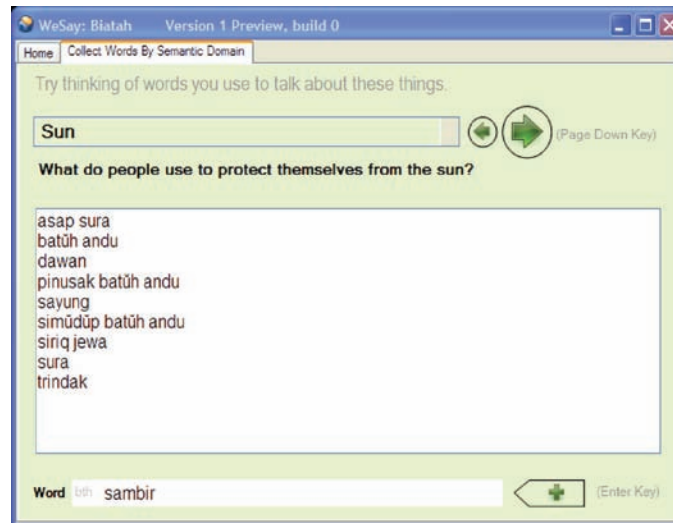


FIGURE 3

He does this by trying to answer questions that are designed to get people thinking about the domain. The user types as many words as he can think of for a question and then moves on to the next question. The words are associated with the domain itself, not the question; so as the user moves between questions, the words for that domain remain. When

the user goes to a new domain, the set of words is cleared and any new words become associated with this new domain. If the user needs to make changes to a word or remove it from the domain, he can select the word and it is brought back down into the editing area.

When the advisor is satisfied that the user understands the task and is confident with the tool, she can let him go until he has completed the task. Since version 4 of Ron Moe's Dictionary Development Process contains almost 1800 domains, this task could take a single person a few months.

When the user has gathered words from each domain, he returns to the advisor who can now set him up with a new task. Let us say the advisor decides the next step is to gloss into English all the words that have been collected. She brings up the configuration tool again and, this time, turns on the "Add Meaning" task. The user is finished with collecting words by semantic domain, so that task can be turned off if the advisor judges that the user is unlikely to go back to collecting more words that way.

The advisor may also choose to turn on the "Dictionary" task, which lets the user search for entries and edit them in place, add words, and delete them. Our advisor decides to have this task only display the meaning (gloss) field and the semantic domains of each sense (Figure 4), since any other fields would just be noise to the user at this point.

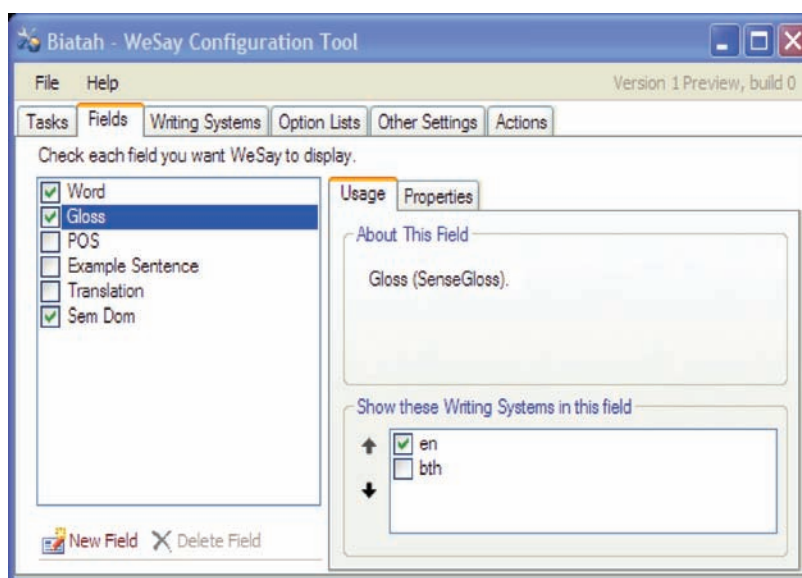


FIGURE 4

Now WeSay shows the "Dictionary" task, the "Actions" task and the "Add Meanings" task.

WeSay offers efficiency gains through task orientation; however, this does not limit the user to a single chance to do each task. There are two ways that a user can return to a task, after completing the bulk of the work in that area. Firstly, if the "Dictionary" task is



enabled, the user is free to add and edit information without using any of the more specific task-tools. Secondly, the dashboard that serves as the home screen for WeSay, enabled tasks indicate to the user when more work has piled up in a task. For example, if the user has added meanings for all the words in the dictionary, but later adds some new words, the “Add Meanings” task will display the number of new words that lack meanings (Figure 5). The number of items to be done in that task is shown in a graphic suggestive of an “in tray”. This invites the user, over time, to return to tasks, keeping all of the entries complete.



FIGURE 5

Once again, the advisor trains the user for the work the user is about to do. Using the “Add Meanings” task, he can now work his way through the dictionary, word by word, adding a gloss to those entries that lack glosses.

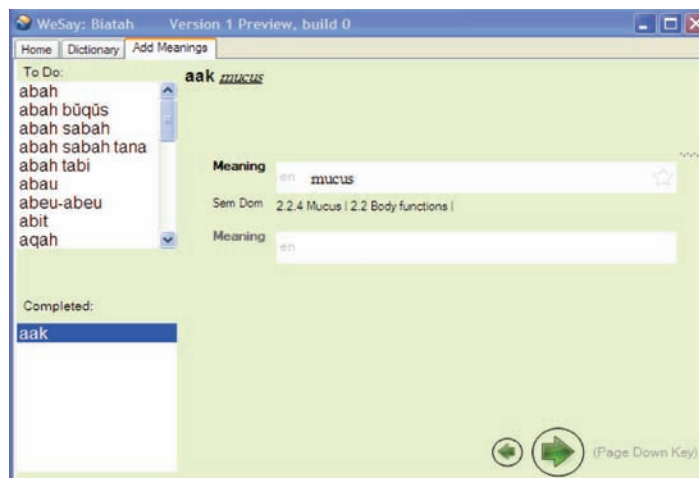


FIGURE 6

So what is the “Dictionary” task good for? Let us say that one evening, while sitting around the fire, the user hears someone use the word ‘dog’ in a way that he does not think is in the dictionary. Back home, he fires up the laptop and goes to the dictionary task. Before he can add a sense, he needs to find the existing entry for ‘dog’. This may be harder than one would think.

Whereas many people take spelling for granted, in many of these situations, orthographies are still in flux and people may not have had a lot of experience writing their own language. This can make finding words difficult while providing ample opportunity for accidentally entering the same word into the system with multiple spellings. To help with this, WeSay suggests words that approximately match<sup>6</sup> what the user types.

Once the user finds the word, he adds a new sense by typing the new meaning in the blank space for the data. Just in case the user is not sure what will happen when he types in a new word, the preview at the top of the screen displays a blank to show where this new meaning will fit in the dictionary entry when it is printed (Figure 7).

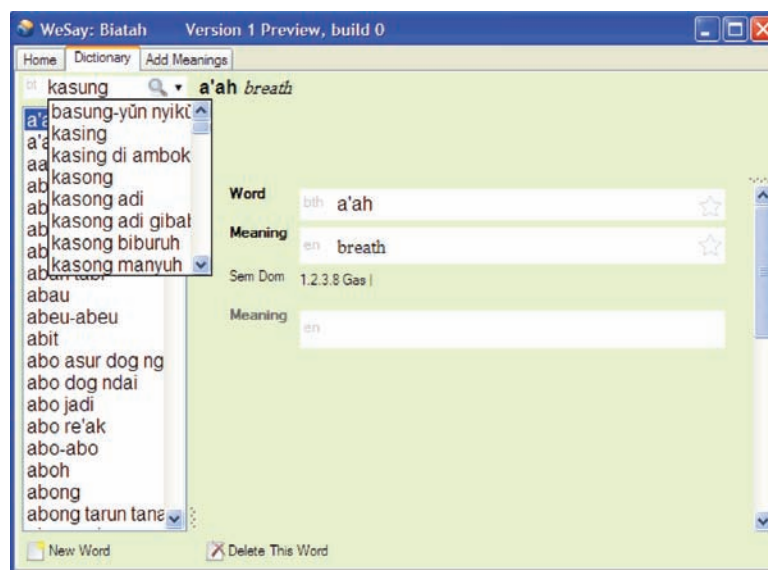


FIGURE 7

When the user has completed the “Add Meanings” task, he returns to the advisor who now enables the “Add Examples” task, and turns on the example field. Each time they come back ready to work on another task, the advisor progressively gives more training and unveils more features.

<sup>6</sup> WeSay uses a modified “edit distance” algorithm for this matching. While it works reasonably well in a wide variety of languages, we may provide the ability to tune it for particular languages in the future.

At any point in the process, the user can print a report that shows his progress or print a copy of the work he has done so far to give to community leaders and members (Figure 8).



FIGURE 8

**6. CONCLUSION.** We believe that indigenous language communities need tools designed from the ground up, for their needs. We cannot just take software designed for expatriate linguists, strip down menus, and pretend that a product is now easy to use. By removing many of the obstacles that speakers of under-resourced languages face, WeSay enables people from under-resourced language communities to create dictionaries for themselves with minimal training on inexpensive, rugged, low-power hardware that can be used within their community. Thus WeSay can play an important role in enabling community members to be active contributors in the process of dictionary creation.

#### REFERENCES

- BUTLER, LYNNIKA and HEATHER VAN VOLKINBURG. 2007. Fieldworks Language Explorer (FLEx). Language documentation & conservation 1:1. <http://nflrc.hawaii.edu/ldc/June2007/techreviews/butler.html>
- COOPER, ALAN. 2004. The inmates are running the asylum. Indianapolis: Sams Publishing.
- CORRIS, MIRIAM, CHRISTOPHER MANNING, SUSAN POETSCH, and JANE SIMPSON. 2002. Dictionaries and endangered languages. In David Bradley and Maya Bradley (eds.), Language endangerment and language maintenance. London: RoutledgeCurzon: 329-347. <http://nlp.stanford.edu/pubs/eldic.ps>
- COWARD, DAVID E. and CHARLES E. GRIMES. 1995. Making dictionaries: a guide to lexicography and the Multi-Dictionary Formatter (Version 1.0). Waxhaw: Summer Institute of Linguistics [http://www.sil.org/computing/shoebox/MDF\\_2000.pdf](http://www.sil.org/computing/shoebox/MDF_2000.pdf)
- FIELD LINGUIST'S TOOLBOX. <http://www.sil.org/computing/toolbox>
- HOSKEN, MARTIN. 2006. Lexicon Interchange Format: A description. [http://lift-standard.googlecode.com/files/lift\\_10.pdf](http://lift-standard.googlecode.com/files/lift_10.pdf)
- JOFFE, DAVID and GILLES-MAURICE DE SCHRYVER. 2004. TshwaneLex – A state-of-the-art dictionary compilation program. In G. Williams & S. Vessier (eds.). 2004. Proceedings of the eleventh EURALEX international congress, EURALEX 2004, Lorient, France, July 6-10, 2004: 99–104. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud. <http://tshwanedje.com/publications/euralex2004-TL.pdf>
- MOE, RON. 2001. Lexicography and mass production. Notes on linguistics 4:3. [ftp://ftp.sil.org/software/win/ddp/doc/ddp4\\_nlx\\_article.doc](ftp://ftp.sil.org/software/win/ddp/doc/ddp4_nlx_article.doc).
- .MOE, RON. 2006. Dictionary development process. SIL International. <http://www.sil.org/computing/ddp>.
- ONE LAPTOP PER CHILD. <http://www.laptop.org>.
- SHORE, SUSAN and RENÉ VAN DEN BERG. 2006. A new mass elicitation technique: The dictionary development program. Paper presented at tenth international conference on Austronesian linguistics, January 17-20, in Puerto Princesa City, Palawan, Philippines. [http://www.sil.org/asia/philippines/ical/papers/vandenberg-mass\\_elicitation\\_dictionary.pdf](http://www.sil.org/asia/philippines/ical/papers/vandenberg-mass_elicitation_dictionary.pdf).
- SIL INTERNATIONAL. SIL FieldWorks Language Explorer features: bulk edit. <http://www.sil.org/computing/fieldworks/flex/bulkedit.html>

Eric Albright  
[eric\\_albright@sil.org](mailto:eric_albright@sil.org)

John Hatton  
[john\\_hatton@sil.org](mailto:john_hatton@sil.org)

# 11

## A Formosan Multimedia Dictionary Designed Via a Participatory Process

Meng-Chien Yang, Hsin-Ta Chou, Huey-Shiuan Guo, Gia-Pyng Chen

*Providence University*

Digital archiving is important work for an endangered language, because if an endangered language disappears, associated cultural assets will disappear altogether. Several digital archiving projects are being conducted in Taiwan. Many tribal teachers are now involved in these projects. Based on the needs of these tribal teachers, this paper presents an easy-to-use system for digitally archiving Formosan Languages. The proposed approach takes advantage of the Internet and the newly launched Web 2.0 sharing platform. This chapter gives details of the development and structure of the online dictionary system. Currently, several archiving projects in Taiwan are using this system to teach tribal teachers how to develop their own language resources and online dictionaries.

**1. INTRODUCTION**<sup>1</sup>. Developing dictionaries for endangered languages is a long and complex process. Although it is easy to collect large archival databases of endangered languages, the purpose and how to best use these archives is sometimes unclear. The most frequently asked question is how the documentation benefits native speakers of the language (Eisenlohr 2004). This question can be addressed simply by creating a shared language resource from these archives. However, implementing this shared language system is a very complicated and difficult task. The Internet is probably the best vehicle for developing shareable language resources. The well-established Lexique Pro, developed by the Summer Institute of Linguistics (SIL), can transform a digital archive into a dynamic dictionary with hyperlinks that can be published on websites. Lexique Pro is a useful tool for field linguists developing shareable language resources. Conversely, mastering Lexique Pro requires both the skills associated with field linguistics and computer technology. Native speakers have recently become increasingly devoted to saving their own languages. Although indigenous peoples often possess considerable knowledge of their own languages, very few have the necessary computer and linguistics skills. Hence, their knowledge can be lost if it is not transformed into shareable digital archives. A process that assists indigenous peoples in creating shareable Internet language resources would be valuable. This project was motivated by the enthusiasm and needs of tribal teachers. The work attempts to create a platform for Formosan tribal language teachers to create their own shareable Internet dictionaries.

Research in endangered language documentation was first recognized as a separate field by Himmelmann (1998). Many computer software tools were designed for field linguists to do the documentation work. The tools available for digital archiving were well described

---

<sup>1</sup> The authors would like to thank Providence University and ELDP, SOAS, University of London for financially supporting this research under contract no. **94 11100 B10**. We also appreciate the editorial services of Dr. Gerald Rau.

in Bird and Simons's work (Bird and Simons 2003a, 2003b). Bird and Huang proposed a platform for language sharing and exchange (Bird, Simons and Huang 2001). Several studies report their collection of endangered language documentation (Lublinskaya and Sherstinova 2002, Psutka, J., et al. 2002, Johnson 2004). The use of the digital archiving for language learning was reported in Csató and Nathan's works (Csató and Nathan 2003, Nathan 2004). In Taiwan, there are several important digital archiving projects (Zeitoun et al. 2003, Zeitoun and Yu 2005). Recently, the research team doing Yami documentation has been developing an e-learning platform and model for Internet e-learning (Rau and Dong 2006, Rau and Yang 2006, Rau, Yang and Dong 2006, Yang and Rau 2005).

This chapter presents a novel web-based system that allows Formosan tribal teachers to create shareable dictionaries of their languages. Formosan tribal teachers are typically the elders, who have little experience using computers. Some tribal teachers are school-teachers or clergy residing in a community. As the government is promoting the use of the Internet and web applications, many tribal teachers may have participated in seminars introducing the Internet and have experience using the Internet. This chapter presents a simple and useful web-based system that allows tribal teachers to create their own web-publishable dictionaries.

This project capitalizes on the openness and freedom of the Internet to design an environment in which the tribal teachers can create and share their languages. The system is based on a participatory process associated with new generation web applications, such as Web 2.0 (Treese 2006). The design combines field experience and IT technologies to create an online environment for developing Formosan dictionaries. This collection of dictionaries can be utilized as a resource for creating CALL systems for teaching Formosan languages (Fujii et. al. 2000, Ward and van Genabith 2003).

The remainder of this chapter is organized as follows. Section 2 includes two case studies from the Yami documentation project. These two cases illustrate some difficulties the indigenous tribal teachers faced in building their own language documentation projects. Section 3 describes the design of the language resource use and editing environment to develop a Formosan dictionary system with user-friendly interface and simplified annotation tools. Section 4 gives a detailed description of this system. Section 5 presents the conclusions.

**2. INITIATIVES AND BACKGROUND OF THE PROJECT.** This project is motivated by the experience and results of a project sponsored by the Endangered Language Documentation Program (ELDP) to document the Yami language (see Rau and Yang in this volume for further details). A language revitalization seminar was held on Orchid Island in 2006. Several tribal teachers were recruited to collect the Yami corpus. To help these teachers document the collected Yami corpus, a seminar was held on how to use Toolbox, a software package for language documentation developed by SIL. The teachers showed great enthusiasm for documenting the Yami language; however the teachers had difficulties using Toolbox. The next section describes the Toolbox training seminar and the difficulties encountered.

The Yami consultants on the research team contributed many corpora in various formats such as audio tapes and video tapes. To document these corpora, the research team spent considerable time digitizing and annotating these analog data. These experiences are described in Section 2.2.

**2.1 THE TRAINING SEMINAR IN USING TOOLBOX ON ORCHID ISLAND.** A Toolbox training seminar was held on April 1 and 2, 2006 at Lanyu High School, Orchid Island. The seminar topics were lexical annotation, glossary creation, word and phrase compilation using Toolbox and producing a digital dictionary. Eight tribal teachers participated in this seminar. Some teachers were teaching the Yami language at local schools. Therefore, the aim of the workshop was to teach the Yami tribal teachers how to construct dictionaries. A sample dictionary previously designed by the research team was provided to participants.

Toolbox operations were introduced at the beginning of the class. Research team members created a set of snapshot steps so that these Yami tribal teachers could quickly create a simple version of the Yami dictionary (Figure 1). However, the research team found that the class members were confused by the complex settings in Toolbox and its English interface. Class progress was far behind schedule at the end of the seminar.

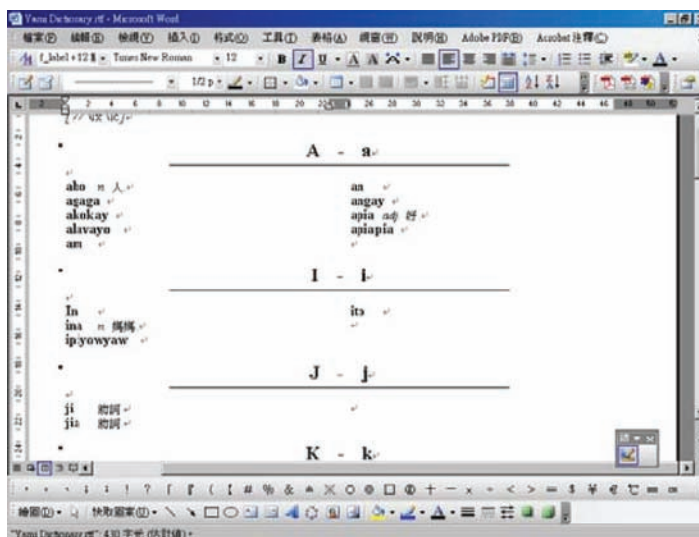


FIGURE 1. Toolbox export dictionary file

To prepare for the introductory seminar, the project team created a Chinese manual with the following content:

1. How to download Toolbox software
2. Basic operations of Toolbox
3. How to build the corpus using Toolbox

Tribal teachers had three major difficulties using Toolbox after the seminar. First, all processes, explanations and interfaces are in English, which clearly presented difficulties for these teachers. Second, Toolbox functions are numerous and diverse. Toolbox users need to define many items when a corpus is created. Finally, the tags and marks of Toolbox are user-oriented and cannot be shared with other users.

Similarly, problems were encountered when instructing tribal teachers how to use Shoebox and Lexique Pro. Lexique Pro software was specially designed for creating shareable language resources. Although Chinese translations of the manuals of these software programs were created and placed on the project web site, [http://yamiproject.cs.pu.edu.tw/yami/yami\\_ch/link.htm](http://yamiproject.cs.pu.edu.tw/yami/yami_ch/link.htm), these difficulties remained. Tribal teachers could not use these manuals to help them develop their own language resources. To assist these tribal teachers in developing their own language corpora, problems were analyzed to find a solution based on the local culture and the abilities of the tribal teachers.

**2.2 CONTRIBUTIONS FROM THE TRIBAL TEACHERS .** The Yami documentation project invited the local Yami consultants to contribute their own language resources. One of our consultants had 101 audiotapes, recorded over three decades. These audiotapes contain many recordings of folk songs, ceremonies, special activities, and teaching from elders. However digitizing these tapes without losing contents was a challenging task, as the quality of the tapes had deteriorated. Some tapes had bad tracks that could not be digitized. Therefore, one staff member manually transformed each track of these tapes into digital data. Transforming legacy language resources into new digital data is a very common practice. We speculate that many such audiotapes exist. These audiotapes must be preserved, organized and transformed into digitally archived data.

**3. PROJECT RATIONALE .** This project explored possible solutions to, and techniques available for creating a shareable language resource for the Formosan languages. The advantage of the Internet was the main consideration in the design of the shared resources. In addition, this project adopted the design of the Web 2.0 platform (Millard and Ross 2006, Treese 2006). Figure 2 shows the format for the participatory process. The objective was to create a web-based online Formosan dictionary editing system. The system was designed as a shareable and easy-to-use platform.



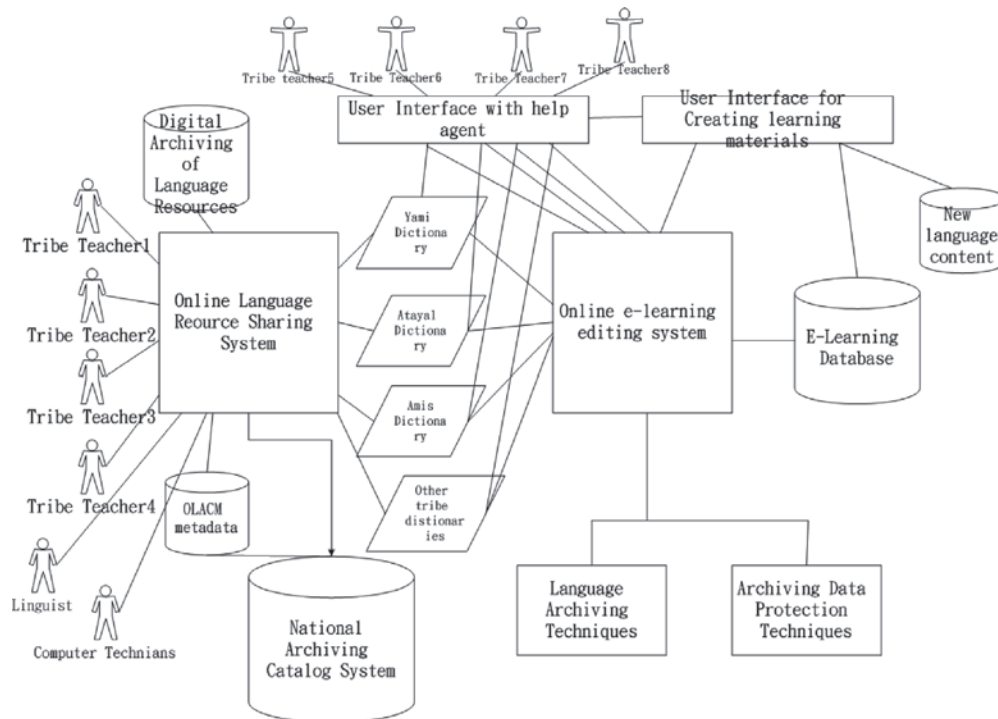


FIGURE 2. Diagram of the system based on the Participatory Process

In this system, the tribal teachers can perform three different tasks:

1. **Collect language resources and annotations:** The online language resource sharing system is designed as the entrance for tribal teachers to collect data. After the collected language resources have the OLACM metadata added, then they can be transformed directly into an online tribal dictionary (Bird, Simons and Huang 2001).
2. **Create online learning materials:** Tribal teachers can create and design their own e-learning materials using the language resources collected.
3. **Share and searching the proper language resources:** The tribal teachers can electronically search and share their language resources.

We hope the proposed system can produce an Internet environment in which all participating teachers can post their language collections and can produce their own language learning materials. This on-line environment would establish a virtual community among the tribal teachers. Based on Web 2.0, this environment should foster sharing and collaborative activities among the tribal teachers of the same Formosan language. Moreover the

system can function as an intelligent diary (de Silva et al. 2007). This on-line environment is being developed using coarse-to-fine and bottom-up strategies. Development is currently at a very early stage and focuses on the following two components.

Interface Design for the tribal language teachers and elders: Most tribal language teachers had not previously taken any computer-related training courses and tend to avoid using computers. Therefore, the application environment must be simple, easy to use and facilitate ‘safe exploration’ for tribal teachers. This project is not a trivial or simple project and must undergo several cycles of evaluation and refinement.

A localized and simplified version of Toolbox for creating the language resources: A simplified version of language documentation tools must be created to record the basic and important lexical items and the collected language words and phrases. This software can hopefully be extended to enable data exchange with databases created in Toolbox.

This online Formosan dictionary editing system, called “Taiwan Austronesian Language Digital Archiving System at Providence University” (TALDAS-PU) can be found at <http://dicts.cs.pu.edu.tw/ada/>. This system allows tribal teachers to enter their new words and eventually produce dictionary-style output.

**4. SYSTEM DEVELOPMENT.** A digital archiving system (TALDAS-PU) is currently being developed for the tribal teachers based on the design described in Section 3. Figure 3 shows the architecture of the system. This system includes a web server for developing server site programs and a database management tool for creating the tables for the digital archiving system.

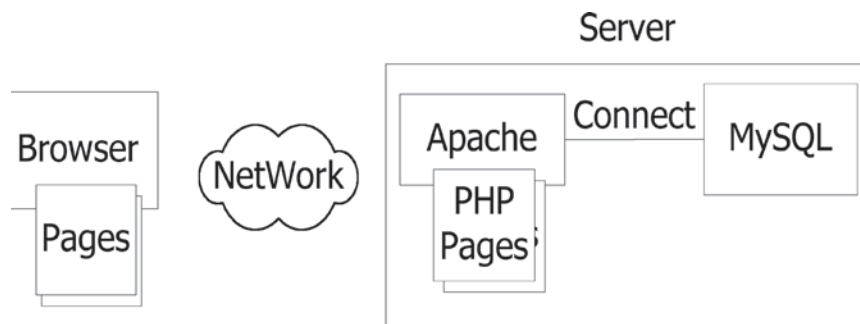


FIGURE 3. System Architecture

The whole system is described as a web-based application. Figure 4 shows the structure and modules of site web pages.

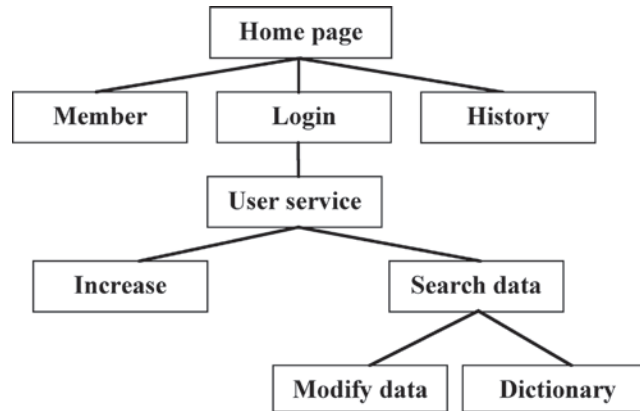


FIGURE 4. Structure of Dictionary System

Users must register when they first log in to the system. Users can then enter service modules and create their own language resources by adding new words to the database, searching the existing word entries or modifying the word entries. Additionally, the system records the action history of the users in the history module. The history data can then be analyzed to give the users proper guidance and assistance.

In each module (Figure 4), several functional sub-modules are created under the block modules:

1.	User registration	This sub-module allows users to register their personal information including name, email address, account, password, language, tribe, job and language used.
2.	Login interface:	This sub-module consists of a user interface for entering the system.
3.	Adding a new entry	This module allows users to add a new word entry.
4.	Data search module	This module provides search functions for the dictionary.
5.	Modifying existing entries	This module allows users to modify existing language data.
6.	Dictionary output module	This module allows users to output the language dictionary.
7.	History module	This module records history data when members add or modify data.

The database for TALDAS-PU consists of five relational tables: User, Language, Vocabulary, DeceX and History. Figure 5 shows the data flow diagram and the program modules of the system. From the main page <Index.php>, a registered user of this system can click the [login] button to login into the web site and first-time users to click the [register] button to register as a user.

In <main.php>, users can add a new entry or search the data of the Formosan languages. When a user clicks the [increase-language-data] button, the web links to <input.php> and the user can create a new language entry. If the user clicks the [search-language-data] button, the page <search3.php> opens. If a user selects the search option and clicks the [search] button, then <result2.php> is displayed and the results of the search will be shown on the page. A user can preview the dictionary by clicking the [Dictionary-Preview] button in <result2.php>. The dictionary is shown on the page <re1.php>. If a user wishes to modify the existing data, clicking the [modify-data] hyperlink opens a new page <modify.php> that contains the data that the user wants to modify. After the user modifies the data and clicks the [modify] button, <modifydata.php> is called to check the correctness of the modified data.

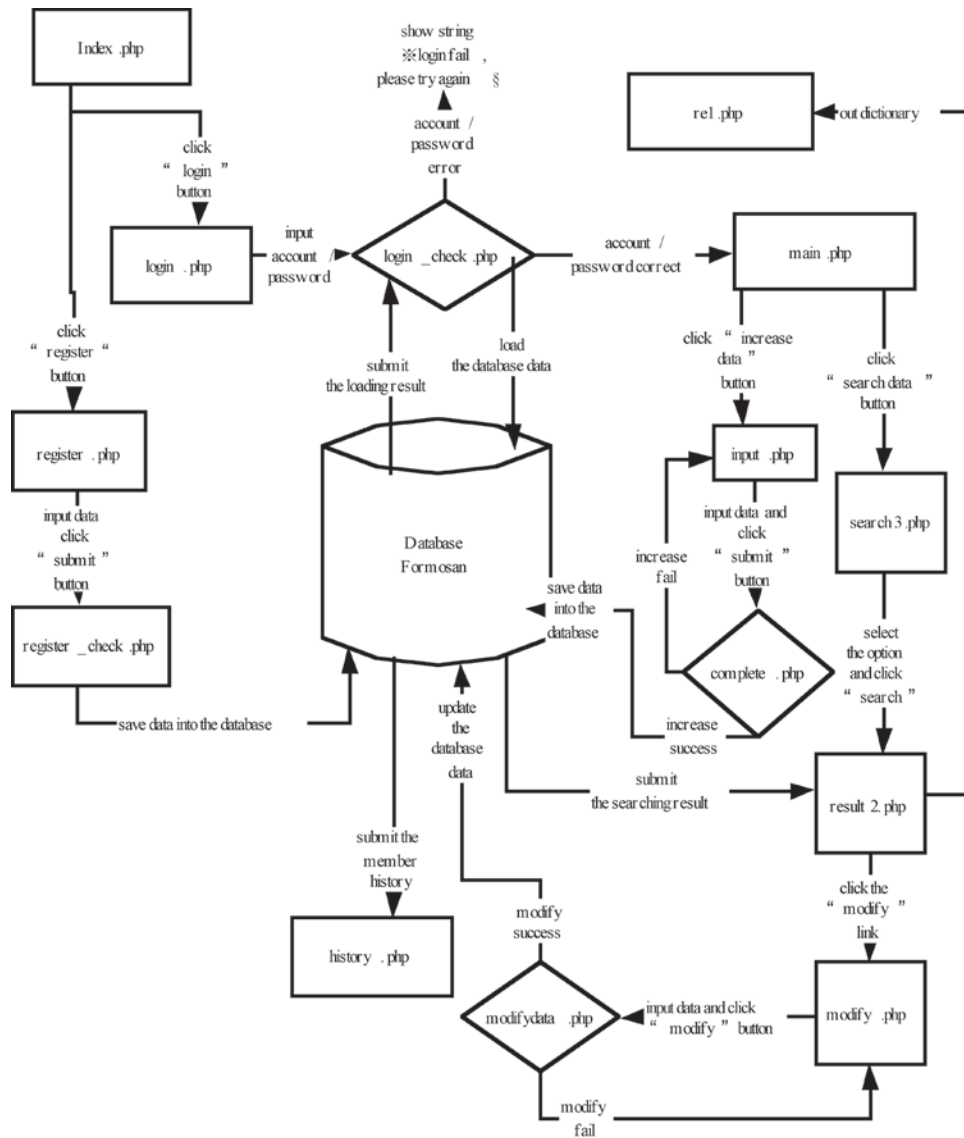


FIGURE 5. System Data Flow Diagram

**4.1 INTERFACE DESCRIPTION.** This section describes the interfaces and functions of the system. The system web site is located on the digital archiving project web site. Interfaces were developed based on studies of tribal teacher needs and the framework of the shareable online dictionary. The important features of the interface are described as follows.

*User Registration Page:* When a user clicks the [Register] button, the User Registration page is accessed (Figure 6). This page has the following eight fields. The fields with an \* mark must be filled in.

- a. Full Name: the user's name
- b. E-mail Address: the user's e-mail address
- c. Username: an identity for the user, defined by the user
- d. Password: the password for entering the system
- e. Language/Dialect: the main Taiwan Austronesian Languages are already registered in the system. The system currently holds 40 Formosan languages with sample words and phrases
- f. Tribe/Location: the tribe or address of a user
- g. Occupation: user occupation, and
- h. Language Use: The situation of the language use such as day-to-day use or use only at work.

FIGURE 6. The User Registration Page

*User Service Page:* Once a user logs in successfully, the User Service interface appears as shown (Fig. 7). Via this page, a user can select one of the two system functions, create a new dictionary or search a specific dictionary.

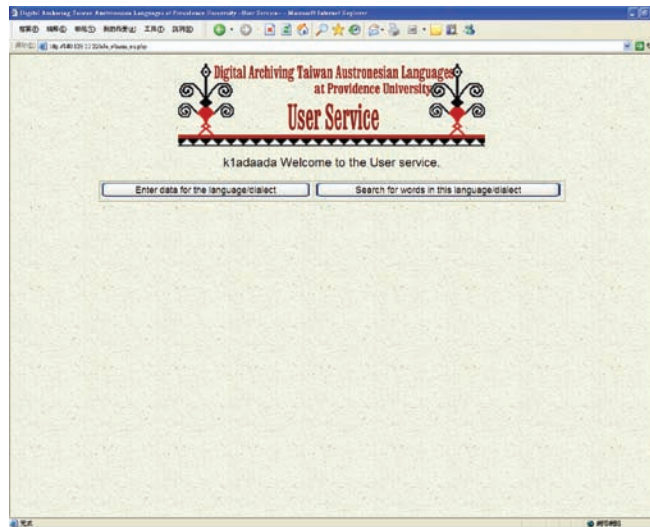


Figure 7. The User Service Page

*Data Entry:* The data entry page (Fig. 8) allows a user, a tribal teacher, to enter a word or a phrase following the steps shown on the page.



FIGURE 8. The Data Entry Page

The design of these fields is based on studies of tribal teachers' needs. This page quickly collects a large number of words with distinctive annotations. This page has fields to input language/dialect data as follows:

- a. Data status: identifies whether the data have been opened. If the field value is “edited”, then the data are being edited and can only be searched or modified by the editor. Other users cannot search or modify the data. If the value is “protected”, then the editor and other users only can search for the data but cannot modify it. If the value is “open”, then the data are open to the public, and all users can search or modify them.
- b. Language/Dialect: records the language/dialect of a word a user wants to add
- c. Entry: a new entry in a language/dialect
- d. Upload Sound of Entry: this functionality allows a user to upload a voice file
- e. Upload Graphic of Entry: this field is for uploading an image or graphics concerning an entry
- f. Root: records the root form of a word
- g. Variant: records word variants
- h. Chinese Definitions and Examples: an extensible field for Chinese explanations and sample sentences.

*Search for Language/Dialect Data:* A user can search any Formosan language/dialect from a dropdown list. Figure 9 shows the search page.

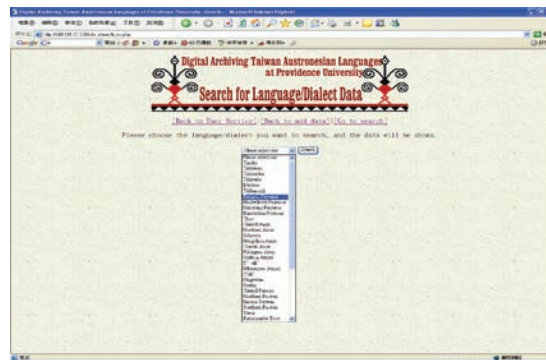


FIGURE 9. The Language Search Page



Figure 10 shows sample search results.

Serial No.	Spoken Form	Spoken Form	Spoken Form	Chinese characters	Example	Spoken Form	Chinese translation of the example	More
1								
2		baki	baku	祖父	Baki o lama tama mu.		祖父是爸爸的爸爸。	More
3		bubu	bubu	祖母	Bubu misu ka isu.		祖母是奶奶。	More
4		buwax		白蟻	Wini sahuq buwax ka gaga.		那是群白蟻的窩。	More
5		dgiyaq		山	Kumay o mniq kaka dgiyaq.		那是我的山。	More
6		embiyax		祖母	Embiyax su hug?		你祖母嗎?	More
7		erut		孩子	Erut sapah mu o snatu xiluy.		那是我的孩子。	More
8		gikus		絲	Gikus ka ga ruwan ubang.		那是我的絲。	More
9		hakaw ibay		神話	Hakaw ibay ka gaga.		那是一個神話。	More
10		hngkawas		菜	Kingal hngkawas o mtgapat karat.		那是一盤菜。	More
11		huling		舞	Mla bi kmlewa sapah ka huling nil.		那是我的舞。	More
12		idaw		甜	Shiya bi uqun ka idaw nil.		那是我的甜。	More
13		lya		不	Iya gauqi ka musa mates.		那不是。	More
14		kacing		水	Kacing balay ka nil.		那是我的水。	More
15		kndadax		出口	Maxal dha tuki ka kndadax ku.		那是我的出口。	More
16		kowbu		味道	Cesagan ni tqian ka kowbu gaga.		那是我的味道。	More
17		kuwar		船	Faw bi kwar ka kumar.		那是我的船。	More

FIGURE 10. Results from a Dictionary Search

Online Dictionary: A user can preview the dictionary for a selected language/dialect by clicking the dictionary-preview button. The online dictionary pops up in a separate window (Fig. 11).

1								
2	baki	祖父	Baki o lama tama mu.	祖父是爸爸的爸爸。				
3	bubu	祖母	Bubu misu ka isu.	祖母是奶奶。				
4	buwax	白蟻	Wini sahuq buwax ka gaga.	那是群白蟻的窩。				
5	dgiyaq	山	Kumay o mniq kaka dgiyaq.	那是我的山。				
6	embiyax	祖母	Embiyax su hug?	你祖母嗎?				
7	erut	孩子	Erut sapah mu o snatu xiluy.	那是我的孩子。				
8	gikus	絲	Gikus ka ga ruwan ubang.	那是我的絲。				
9	hakaw ibay	神話	Hakaw ibay ka gaga.	那是一個神話。				
10	hngkawas	菜	Kingal hngkawas o mtgapat karat.	那是一盤菜。				
11	huling	舞	Mla bi kmlewa sapah ka huling nil.	那是我的舞。				
12	idaw	甜	Shiya bi uqun ka idaw nil.	那是我的甜。				
13	lya	不	Iya gauqi ka musa mates.	那不是。				
14	kacing	水	Kacing balay ka nil.	那是我的水。				
15	kndadax	出口	Maxal dha tuki ka kndadax ku.	那是我的出口。				
16	kowbu	味道	Cesagan ni tqian ka kowbu gaga.	那是我的味道。				

FIGURE 11. Dictionary Preview

*Modify Data*: To modify a word in an online dictionary, a user clicks the [Modify-Data] button to open a new “modify” page for this word. Figure 12 shows this “modify” page.

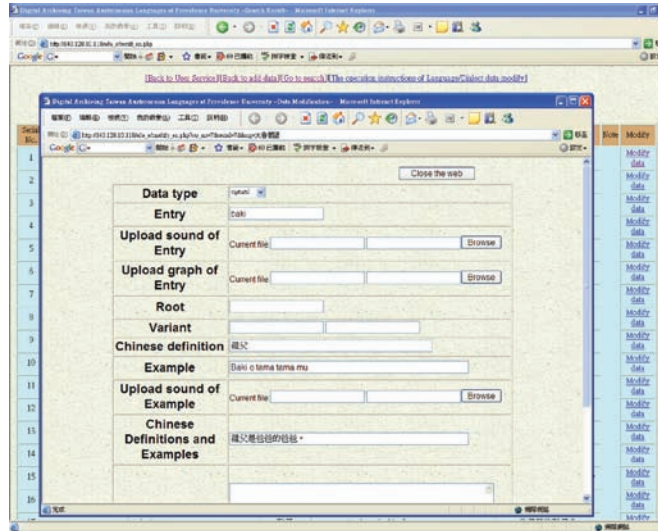


FIGURE 12. The Data Modification Page

*Log of Action History*: A log function records user activity. If a user adds a new word or modifies a word successfully, the system records this action in the database. Figure 13 shows a user action log. This information is used to analyze system usage.

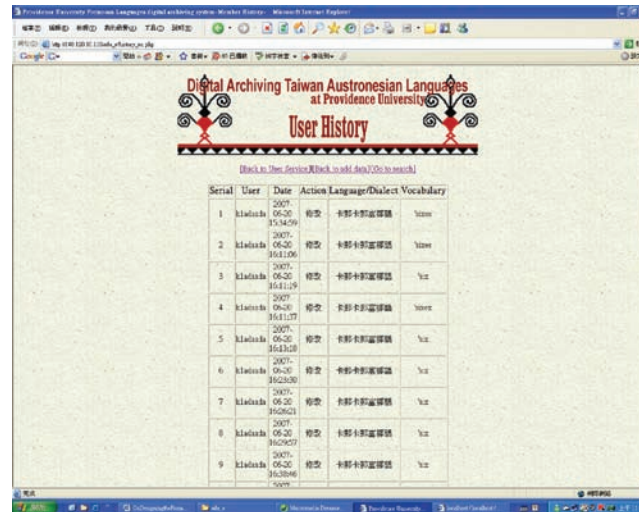


FIGURE 13. The Action Log of User History

**5. CONCLUSIONS.** This chapter has described the design and implementation of an attempt to create an online Formosan language resource sharing and editing system. The design idea is based on a participatory process for sharing language resources. Development of “Taiwan Austronesian Languages Digital Archiving System at Providence University” (TALDAS-PU) is also reported. The system is a web-based system that can be accessed by any web-based platform and viewed using a browser. Therefore, the system is a feasible platform for online language documentation on a very compact computer, such as an OLPC (<http://laptop.org/>).

Future work will develop and finish all components. In the next stage, this project will focus on how to use the Web 2.0 platform to create proper tools for sharing the language resources.

## REFERENCES

- BIRD, STEVEN, GARY SIMONS, and CHU-REN HUANG. 2001. The open language archives community and Asian language resources. *NLPRS 2001*: 31-38.
- BIRD, STEVEN and GARY SIMONS. 2003a. Extending Dublin Core Metadata to support description and discovery of language resources. *Computers and Humanities* 37: 378-388.
- BIRD, STEVEN and GARY SIMONS. 2003b. Seven dimensions of portability for language documentation and description. *Language* 93: 557-582.
- CSATÓ, EVA and DAVID NATHAN. 2003. Multimedia and documentation of endangered languages. In Peter Austin (Ed.) *Language Documentation and Description* 1: 73-84. London: SOAS.
- DE SILVA, GAMHEWAGE C., TOSHIHIKO YAMASAKI and KIYOHARU AIZAWA. 2007. An interactive multimedia diary for the home, *IEEE Computer* 40(5): 52-59.
- EISENLOHR, PATRICK. 2004. Language revitalization and new technologies: Cultures of electronic mediation and the refiguring of communities. *Annual Review of Anthropology* 33: 21-45.
- FUJII, SATORU, JUN IWATA, MAYUMI HATTORI, MUTSUMI IJIMA, and TADANORI MIZUNO. 2000. "Web-Call": a language learning support system using Internet. In Proceedings of the Seventh International Conference on Parallel and Distributed systems. 326-331.
- HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-195.
- JOHNSON, HEIDI. 2004. Language documentation and archiving, or how to build a better corpus, In Peter Austin (Ed.) *Language Documentation and Description* 2: 140-154. London: SOAS.
- LUBLINSKAYA, MARINA and TATIANA SHERSTINOVA. 2002. Audio collections of endangered arctic languages in the Russian Federation. In *Text, Speech, and Dialogue: 5<sup>th</sup> International Conference Proceedings* ed. by Petr Sojka, Ivan Kopeček and Karel Pala, 347-356. Springer.
- MILLARD, DAVID E. and MARTIN ROSS. 2006. Web 2.0: Hypertext by any other name? In Proceedings of the seventeenth conference on Hypertext and hypermedia: 27-30. New York:ACM.
- NATHAN, DAVID. 2004. Developing multimedia documentation. In Peter Austin (Ed.) *Language Documentation and Description*. 2: 154-168. London: SOAS.
- PSUTKA, JOSEF, PAVEL IRCHING, JESEF V. PSUTKA, VLASTA RADOVÁ, WILLIAM J. BYRNE, JAN HAJIC, SAMUEL GUSTMAN and BHUVANA RAMABHADRAN. 2002. Automatic transcription of Czech language oral history in the MALACH project: Resources and initial experiments, In *Text, Speech, and Dialogue: 5<sup>th</sup> International Conference Proceedings* ed. by Petr Sojka, Ivan Kopeček and Karel Pala, 253-260. Springer.
- RAU, D. VICTORIA and MENG-CHIEN YANG. 2006. Digital transmission of language and culture rethinking pedagogical models for e-learning, Paper presented at the Joint AAAL and ACLA/CAAL Conference. Montreal, Canada. 6/17-20/2006.
- RAU, D. VICTORIA, MENG-CHIEN YANG and MAA-NEU DONG. 2006. Endangered language documentation and transmission. Paper presented at the 10<sup>th</sup> International Conference on Austronesian Linguistics. Palawan, the Philippines. 1/17-20/2006.

- RAU, D. VICTORIA and MAA-NEU DONG. 2006. *Yami texts with reference grammar and dictionary*, Language and Linguistics. Special Monograph A-10. Taipei: Institute of Linguistics, Academia Sinica.
- RAU, D. VICTORIA. 1995. Yami Vitality. NSC report (NSC84-2411-H-126-001), presented at the Symposium on Language Use and Ethnic Identity, Institute of Ethnology, Academia Sinica 5/16/1995.
- TRESE, WIN. 2006. Web 2.0: Is it really different? *netWorker*. 10(2): 15-17.
- WARD, MONICA and JOSEF VAN GENABITH. 2003. CALL for endangered languages: Challenges and rewards. *Computer Assisted Language Learning* 16(2-3): 233-258.
- WOYCHOWSKY, EDMOND. 2006. *AJAX: Creating web pages with asynchronous JavaScript and XML*. Pearson Education, Inc.
- YANG, MENG-CHIEN and D. VICTORIA RAU. 2005. An integrated framework for archiving, processing, and developing learning materials for an endangered Aboriginal language in Taiwan. In *Proceedings of the 5th Workshop on Asian Language Resources*, 32-39.
- ZEITOUN, ELIZABETH, CHING-HUA YU, and CUI-XIA WENG. 2003. The Formosan language archive: Development of a multimedia tool to salvage the languages and oral traditions of the indigenous tribes of Taiwan, *Oceanic Linguistics*, 42(1):218-232.
- ZEITOUN, ELIZABETH and CHING-HUA YU. 2005. Language analysis and language processing, *Computational Linguistics and Chinese Language Processing*, 10(2): 167-200.

Meng-Chien Yang  
mcyang2@pu.edu.tw

Hsin-Ta Chou  
ada@hinet.net.tw

Huey-Shiuan Guo  
hsguo@pu.edu.tw

Gia-Pyng Chen  
png@pu.edu.tw90

# 12

## Annotating Texts for Language Documentation with Discourse Profiler's Metatagging System

Phil Quick

*SIL International*

This paper introduces a systematic and robust way to annotate (or 'tag') texts with discourse information. To date there has not been a method for annotating texts for language documentation with discourse-text information. This is the first paper to systematically describe the capabilities and the annotating methodology of the *Discourse Profiler's* metatagging system as a means of annotating endangered languages' texts in a *Toolbox* database. Since there is a division of labor between *Toolbox* and *Discourse Profiler*, the *Toolbox* database can be the basis for the archival tasks, whereas the *Discourse Profiler* software is a computer assisted discourse-text analytical tool that mines the *Toolbox* discourse-text annotated database in order to produce two primary capabilities: 1) to create a representative interactive compressed representation or 'map' of the structure and elements of a text, and 2) to quantify texts based on this special metatagging system with an array of sixteen different possible statistical outputs (including both referential distance and topic persistence statistics). Although the main focus of this paper is on the multipurpose annotation system, I will introduce the basics of the *Discourse Profiler* software in order to illustrate the range of analytical possibilities that this annotation system incorporates.

**1. INTRODUCTION.**<sup>1</sup> This paper introduces a systematic and robust way to annotate (or 'tag') texts with discourse information. There are two primary features of the *Discourse Profiler* software (available at [www.discourseprofiler.com](http://www.discourseprofiler.com))<sup>2</sup>. First, it produces an interactive, representative model of a text with a map-like abstraction. A key component of a text's map is the participant tracking which is represented with vertical lines, geometrical shapes for noun phrase types, and colors used to identify either the grammatical role or semantic role. Other vertical lines or 'spans' are used to trace the flow of information in a text parallel to the participant tracking map grid (i.e. span analyses). Second, it permits sixteen different possible statistical outputs, key amongst which are ten different topic continuity statistics that are automatically produced from the annotated text (including both referential distance and topic persistence statistics).

---

<sup>1</sup> I want to thank participants at the International Conference on Austronesian Endangered Language Documentation (June 2007, Taiwan) for their comments and feedback to an earlier version of this paper. I also thank the two anonymous reviewers for their input which has helped to improve this paper. My special thanks to Margaret Florey and D. Victoria Rau for some additional editorial help which goes a long ways toward streamlining my paper. I am responsible however for the presentation and any remaining problems, and I welcome further comments.

<sup>2</sup> *Discourse Profiler* is currently freely available at the website in its current beta 4.4 prerelease version. I welcome input and feedback on its development, including suggestions on the metatagging system. The current plan is to release version 1.0 as shareware.

*Discourse Profiler* has been developed to model and quantify texts based on a special metatagging system. This metatagging system was developed to work with any language, however the particular relevance focused on in this paper is on presenting this as an annotation system that can make a contribution to the documentation of endangered languages. To date there has not been a method for annotating texts for language documentation with discourse-text information.<sup>3</sup>

Although providing some basic information on discourse is clearly useful, the motivation for adding discourse-text information to the annotation of a text for language documentation should be made clear. Discourse-text information clearly adds to the overall richness of what we can learn about a text and its language. It is also the area of language that informs us about lower levels in the hierarchy of a language, especially in the domain of syntax (e.g. word order choices, voice selection criteria, transitivity identification, etc.), as well as intermediate levels such as paragraph structures, episode structures, and how propositional relations make a contribution to understanding the flow and texture of a text.

It is clear that a tagging system for a text should be usable for multiple purposes. A number of linguists have used Microsoft Excel as a means to do topic continuity analysis on various texts of individual languages.<sup>4</sup> The two serious drawbacks of these approaches are that 1) the ‘tagging’ (or annotating) method used has a single purpose and is for all practical purposes useful for only one time or one task, and 2) it is highly laborious. It is clearly more useful to have a metatagging system that has multiple uses because it is more efficient to annotate or tag something one time as opposed to multiple times. Even when the tagging system is laborious, it is more likely that a text will be tagged when the linguist knows there will be multiple uses to that tagging system.

A tagging system also needs to have constraints yet be as flexible as possible. For example, it is more flexible if it can handle multiple theoretical views of syntax. This includes being able to handle encoding all types of clauses (and even clause ‘fragments’). Verbless clauses versus verbal clauses for example need to be differentiated yet to fit within the tagging constraints. The tagging system should not be too difficult to ‘read’, it should be fairly transparent, or at least easily mapped to allow for a simple interpretation process. For texts, it is also important to be able to track individual referents as well as ‘plural’ or ‘mass’ referents, e.g. ‘they’, ‘the children’, ‘trees’, etc. The tagging system should also not

---

<sup>3</sup> For example, it is not mentioned at all in Schultze-Berndt’s excellent 2006 paper on annotating texts Schultze-Berndt (2006) does discuss ‘discourse analysis’, but this is with the meaning of ‘conversation analysis’. *Discourse Profiler* was developed largely to deal with narratives and other lengthy monologues. This does not rule out that some conversation analysis could be done with some changes (possibly minor) in the annotation procedures. Rhetorical Structure Theory deals with some discourse-text information (see Taboada and Mann 2006). However as I understand it, it deals primarily with what I would call propositional relations. There is a developed annotation system for RST, however it appears generally more complex than what would typically be needed for an archival record.

<sup>4</sup> Cliff Olson (pers. comm.) has told me about a tool developed in SIL’s Papua New Guinea Branch. I have heard of other people who have each developed their own custom approaches to doing topic continuity statistics on texts, all of which as far as I know were never used again.

be overloaded with information nor try to capture all information that is in a text. The tagging system should also be easily modified without a serious change in the parameters. It should also be a system that is easily implemented.

The metatagging system developed for the *Discourse Profiler* software fits all of these criteria. The strength of this metatagging system is that it already is multi-purposed for modeling texts and quantifying texts with the *Discourse Profiler* software package, yet can still be adapted, modified, or expanded for other uses. The tags contain a substantial amount of information as will be shown by the wide range of discourse analyses that can be used to model texts visually (with a large range of possibilities, incorporating ideas from Grimes 1975, Longacre 1983, 1996, Givón 1983, 1994, and Quick 1997 among others), and a variety of statistical approaches including a number of topic continuity statistical approaches (e.g. Dooley and Levinsohn 2001, Dryer 1994, and Givón 1983, 1994).

The following section introduces the *Discourse Profiler* software. Although the focus of this paper is on the metatagging system, it is important to understand what kinds of information can be analyzed using a multi-purposed systematic system for annotating a text. The introduction to *Discourse Profiler* demonstrates the robustness of this metatagging system. In Section 3, I introduce the two types of fields that are productive for annotating a text: information type fields and participant tracking fields. In Section 4, I discuss the main clause as the typical unit of description. The following section then briefly examines three features in Toolbox that especially are of help in annotating a text or working with the finished annotated text. In Section 6, the use of *Discourse Profiler* is illustrated through its application with various grammatical categories drawn from the endangered Pendau language (Sulawesi, Indonesia). The conclusion highlights the benefits of *Discourse Profiler* for the documentation of endangered languages and discusses a number of features to be developed in the future.

**2. DISCOURSE PROFILER: A COMPUTER TOOL FOR MODELING AND QUANTIFYING NATURAL LANGUAGE TEXTS.** Linguistic work on language texts has until now lacked a serious software tool that integrates basic linguistic theory and discourse theory. The *Discourse Profiler* software package provides linguists with a tool to analyze the entire context of specific discourse features.<sup>5</sup> These features are summarized here. Practical applications and benefits of using the *Discourse Profiler* program have already been implemented in Quick (2002, 2003, 2005, 2007) for the Pendau language (Sulawesi, Indonesia). Examples of use of the program with Pendau are included in this section.

**2.1 A ROADMAP FOR TEXTS.** A visual model of a text is analogous to a roadmap. Different size cities are represented by an iconic change in size. In a text, different noun phrases can be identified by a different shape, such as using circles and squares to contrast nominative case and accusative case. Colors in roadmaps are often used to contrast geographical features such as blue for water. Colors in *Discourse Profiler* are used to represent semantic or grammatical details such as red for semantic agent, and black for semantic patient. In addition to the basic participant tracking that is displayed in map-like format, syntactic and/or discourse information such as same subject/different subject, event/non-

<sup>5</sup> Earlier beta versions were called *Multilinear Discourse Analysis (MDA)*. The first demonstration version was demonstrated in 1996 (see Quick 1996).



event, and word orders can be traced parallel to the basic text's map display. This allows for analysis of a range of possibilities and the interactive capabilities adds the further help of trying 'what if' easily and rapidly for shifting to various hypotheses and to eliminate or elucidate patterns.

The visual model or map of a text allows the linguist to view fifty clauses and thirty participants in a single view, and therefore clusters of information can be easily compared to the participant tracking. These comparisons and flexible changing of settings (e.g. change the color of subjects from blue to black) allows the linguist to identify possible patterns that are often difficult or impossible to determine when analyzing the actual text.

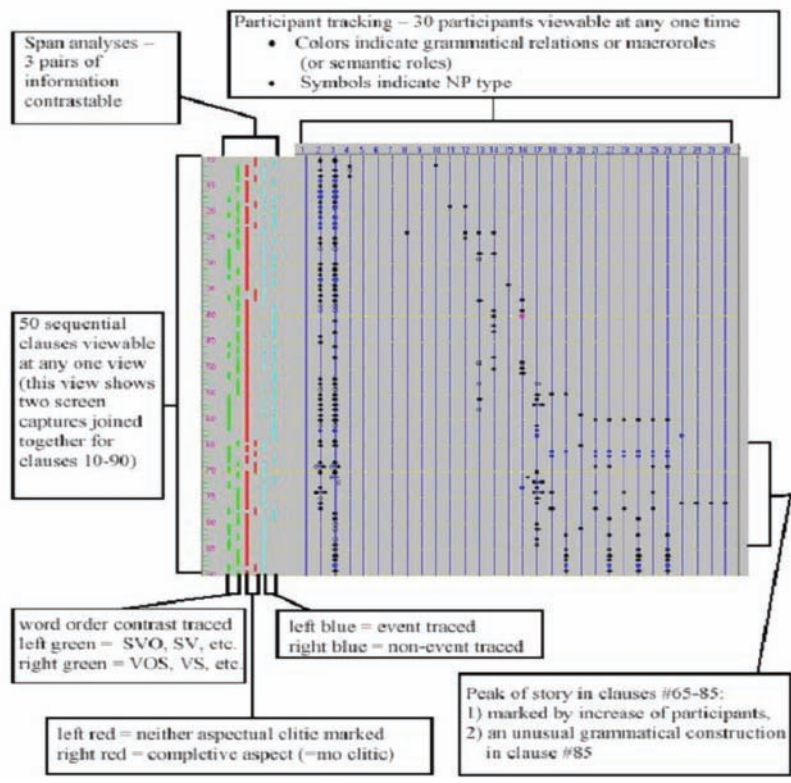


FIGURE 1. Abstract profile of the peak of a Pendau folktale (clauses 10-90) in the *Discourse Profiler* program

With visual modeling of Pendau data, I was able, for example, to compare the span analysis of word order (usually SV or SVO), the occurrence of completive aspect, and the occurrence of certain temporal relators (e.g. *ila uo* ‘after that’). When a clause has a participant change, I could confirm where the boundaries of paragraphs in Pendau occurred most frequently (Quick 2003, 2007).<sup>6</sup>

Figure 1 shows a screen view of clauses 10-90 (top to bottom) of a short recorded text as an abstract visual model or profile. It provides a view of the participant tracking of participants 1-30 (from left to right). Participants are the various geometrical shapes (dots, squares, circles, etc.). This view is a ‘map’ of a text and allows the user to interact with the syntax and discourse in many different ways. The main part of the ‘map’ provides a means for participant tracking. The user assigns different symbols to a NP type, for example a circle might represent a pronoun, and a square might represent a basic NP. Colors can be used to track grammatical relations, e.g. red might indicate a grammatical object, and green might indicate a grammatical subject, or semantic roles (or macroroles) can be contrasted likewise. On the far left are pairs of vertical lines used for span analyses (Grimes 1975). These allow the discourse analyst to track or trace discourse and/or syntactic level information that parallels the participant tracking. For example, event and non-event can be contrasted parallel to different word orders, e.g. contrasting SV/SVO with VOS/VS.

Another well-known literary/discourse feature that occurs in the story profiled in Figure 1 is the gathering of a lot of participants in the peak (see Longacre 1983, 1996). The two main participants are easily identified by participant tracking lines number 2 and 3. Other participants appear fairly randomly until we get near the bottom of this screen view (which is also near the end of the folktale). The zone of turbulence that occurs in a discourse peak can actually be seen visually here in the preceding clauses as the number of participants increases for a number of clauses between 65 and 85. The peak of this narrative has an uncommon grammatical construction (in clause 85) which has not been documented anywhere else in my corpus, but when checked in elicitation it was not considered to be at all unusual.<sup>7</sup>

**2.2 METATAGGING.** The annotation system includes a metatagging system that allows an integrated approach to text analysis and/or text annotation. The entry for data in Discourse Profiler is expedited by utilizing the *Toolbox* software that many linguists are already familiar with. Although the term ‘metatagging’ can generally be thought of as synonymous to ‘annotations’ it really is distinct from the typical annotations that, for example, Schultze-Berndt (2006) describes. This is because the two field types are actually mini databases (note especially the participant tracking fields with its five ‘tiers’). However, since

---

<sup>6</sup> This interaction and visual inspection of the map was only part of the process for identifying paragraph boundaries. There was also some statistical analysis performed on the occurrence and context of the completive aspect *=mo*. Paragraph boundaries may also include changes in location, time, and a change in the setting as well as such things as the beginning of direct speech.

<sup>7</sup> Longacre (1983, 1996; also see Edmondson and Burquest 1992:84-90) describes a number of linguistic signals that indicate a narrative’s peak. This story has at least two of these signals: an increase of participants and an uncommon grammatical construction.

it does fit in generally with the basic notion of annotating at least from the user's perspective, I adopt the term annotation when the focus is on working in *Toolbox*, and use the term *metatagging* when I am focusing on using the *Discourse Profiler* software.

*Toolbox* (and its predecessor *Shoebox*) has become a significant software tool for developing and maintaining lexicons and texts, and for outputting dictionaries. The interlinearizing feature has helped linguists become more productive with its semi-automatic feature of building up a lexicon through the interlinearization process. Although the methodology presented here can be used independent of *Toolbox*, I will assume for ease of discussion that the majority of linguists who will adopt this annotation methodology will at least be using *Toolbox*, and likely will want to use *Discourse Profiler*.

Figure 2 shows a flow diagram of the relationship between *Toolbox* and *Discourse Profiler* and how these annotations are used in each. Data entry of the annotation fields is carried out only in *Toolbox*. The separation of the text database from *Discourse Profiler* is an important one as it allows archiving of plain texts without any interference from *Discourse Profiler*.<sup>8</sup> The flow diagram shows that there are two types of annotation fields: 1) information type fields, and 2) participant tracking fields. These will be further explained

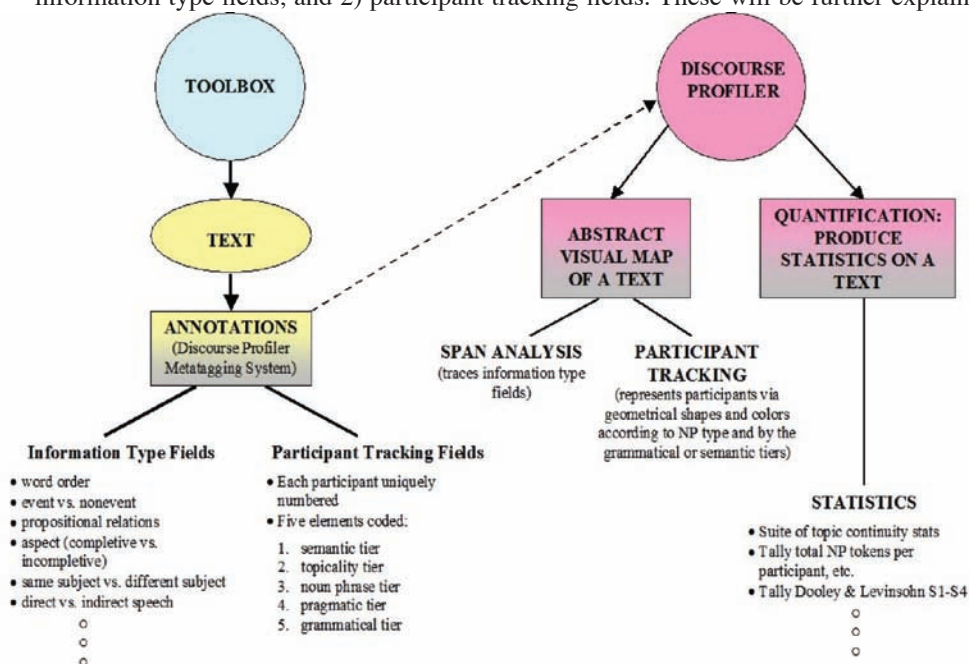


FIGURE 2. Flow Diagram of How the Metatagging System is Used in Toolbox and Discourse Profiler

<sup>8</sup> This also allows other linguistic analysis to be carried out in *Toolbox*, since certain grammatical and discourse information will then be available.

in later sections, but it is important to note that the data entered as annotations here are multi-purposed in order to serve multiple analytical possibilities. After these annotations are entered (also referred to here as ‘metatagging’) along with other language annotations the linguist may be making, the database can be: 1) archived and 2) further analyzed at any time with *Discourse Profiler* or some future software.

The only link between *Discourse Profiler* and *Toolbox* is when *Discourse Profiler* mines these multi-purposed annotated data. This mining only extracts information and does not adjust the *Toolbox* database in any way.

One of the benefits of the *Discourse Profiler* program is that it allows the user to make rapid shifts and/or refine one’s hypothesis or analysis, and to analyze a higher number of texts more efficiently. This flexibility is made possible by the metatagging system and their use within the span analysis settings feature of *Discourse Profiler*. For binary information such as the contrast of event and nonevent, these are simply ‘listed’ in two separate lists so that it is easy to trace the occurrence of each one on separate span ‘lines’ in the map of the text. Information that has more than two items is grouped into two different lists.

For example, in Pendau I list the two contrastive word orders of SV/SVO and VOS/VS (and many other variations which may include obliques and clauses with zero anaphora) into two groups (or lists). These two separate lists of the word orders allow one to work on hypotheses to determine why there is a variation in this word order difference. These groupings are easily changed in these lists so that the span analysis can be compared with other spans and with the text’s participants. Likewise, the settings for participants allow one to reassign the colors and/or shapes easily and quickly. This allows for a great degree of flexibility in increasing the possible analyses. All of these settings can be saved for different analyses of the same text, or using one or more of these settings for as many other texts that will be analyzed. By using the *Discourse Profiler*’s settings there will be nothing in the original *Toolbox* database that will be changed.

**2.3 STATISTICS.** A range of statistical options often used to analyze texts manually is now automated. The time saved can thus be put into analyzing a larger number of texts. The quantification of texts ranges from various topic continuity statistics (e.g. the Givón 1994 approach is different from the Dryer 1994 approach) to basic statistics on the number of noun phrases for each participant tracked.

Statistical analysis was undertaken for the topic continuity of the core arguments in two different transitive constructions in Pendau. This analysis provides evidence that both of these constructions are equally transitive. Active and inverse verbal clause constructions are nearly equal in frequency of occurrence. Discourse topic continuity studies show that the A argument in inverse voice clause constructions is highly topical in this language, and a comparison of the A and P arguments in both active voice and inverse voice clause constructions have a similar profile as expected for transitive clauses. Table 1 illustrates this with one of the texts (Mtext3) from which the statistics were generated with *Discourse Profiler*.<sup>9</sup>

---

<sup>9</sup> The generalizations were gleaned from four texts (Quick 2003, 2005, 2007)

	<i>Inverse Voice</i>	<i>Inverse Voice</i>	<i>Active Voice</i>	<i>Active Voice</i>
RD	P	A	P	A
1-3	70 (70.71%)	65 (82.28%)	24 (53.34%)	79 (89.77%)
>3	29 (29.29%)	14 (17.72%)	21 (46.76%)	9 (10.23%)
Total	99 (100%)	79 (100%)	45 (100%)	88 (100%)

TABLE 1. Referential distance values in Pendau—Mtext3

Topic continuity statistics also suggest that in Pendau previous discourse information is important in the speaker's choice between active voice and inverse voice. A matrix such as shown in Table 2 is produced with the raw data of each text analyzed. The data can then be copied into a text editor for better formatting as illustrated with Table 2. In the case of Table 2, the data from the four texts (Mtexts 1-4) were added together in order to create a final version. Table 2 illustrates the version of topic continuity statistics developed by Dryer (1994).

	Inverse Voice	Active Voice	Total
RD of A lower	60 (38%)	100 (62%)	160 (100%)
RD of A and P same	83 (72%)	33 (28%)	116 (100%)
RD of P lower	100 (60%)	67 (40%)	167 (100%)

TABLE 2. Relative referential distance of As and Ps (Dryer horizontal analysis)  
(All texts combined—Mtexts 1-4)

**2.3.1 GRAPHING.** The statistics produced for each of the four Pendau texts can also be plugged into a graphics program (Figure 3).<sup>10</sup> As typically produced for topic continuity studies, a scatter plot graph is used. For the Pendau data, if the referential distance of the Undergoer is less than the distance for the Actor within the same clause ( $P < A$ ) or the referential distance for the Actor and Undergoer of the same clause is the same ( $P = A$ ), then the inverse voice verbal construction will more often be chosen. But if the referential distance of the Actor is greater than the Undergoer in the same clause ( $A < P$ ), then the active voice

<sup>10</sup> All of the statistics produced are raw data. If the linguist wants to make graphs, then s/he will need to use a graphing program. This usually just involves plugging in the raw numbers, and then producing the graph. All of the matrices produced as raw statistics in *Discourse Profiler* can be copied and entered into a table for better formatting. I have been able to format and graph all of the statistics on Pendau using MS Word.

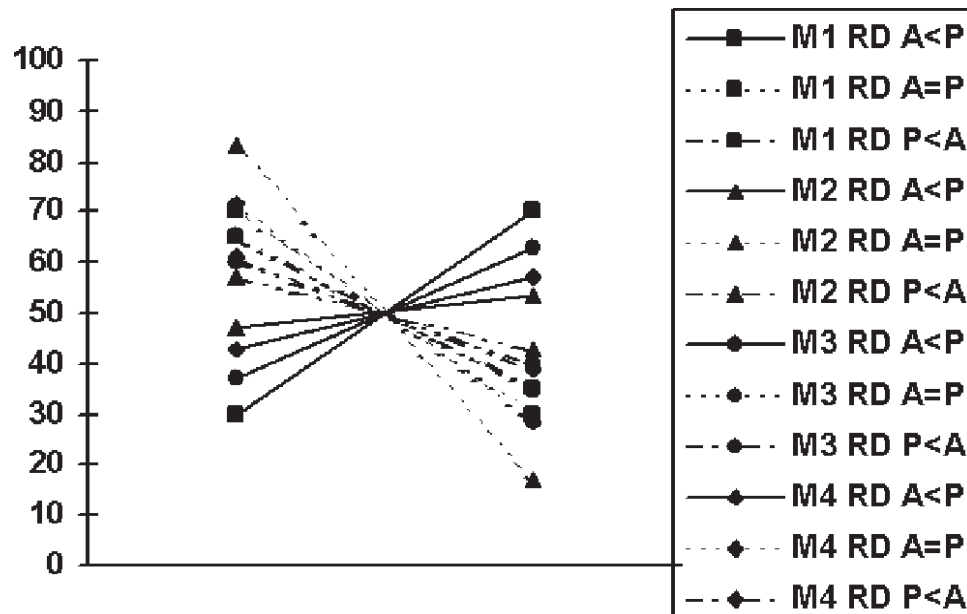


FIGURE 3. Frequency of A and P arguments in active and inverse voice constructions according to whether the A is equal to P in referential distance (RD A=P), the A is less than P in referential distance (RD A<P), or the P is less than A in referential distance (RD P<A); the data are from four Pendau texts, labeled as M1-M4

verbal construction will more often be chosen (see Figure 3). Figure 3 shows there are four areas of significant clustering that occur in the scatter plot graph (roughly in each of the quadrants; M1-M4 refer to four analyzed texts, in which the M reflects the previous beta versions' name for *Discourse Profiler*, which was *Multilinear Discourse Analysis*).

Another basic but potentially highly useful statistical approach is to tally the number of participants according to noun phrase types in various ways. Figure 4 shows one approach to tallying referents of a text with a bar graph. This shows how many occurrences of each participant in a text occur as a different type of noun phrase. For example, in the text quantified in this bar graph it is easy to ascertain that the most common participants are participants one and two (the numbering of the participants is from left to right within each noun phrase type). It is also noted that participant one (red bar on far left of each NP type) appears most frequently as a genitive case pronoun, followed by zero anaphora, and then as a typical noun phrase in absolute case and then in the genitive case. Participant two (brown bar, second from left in each NP type) appears very frequently as a typical noun phrase in the absolute case and very high frequency in zero anaphora. Its other occurrences are quite low.

Two other ways to make similar tallies include how each participant occurs as subject, object, oblique and which type of noun phrase these occur as. Different tallies of a text can be done to compare different cases, participants with relative clauses, where demonstra-

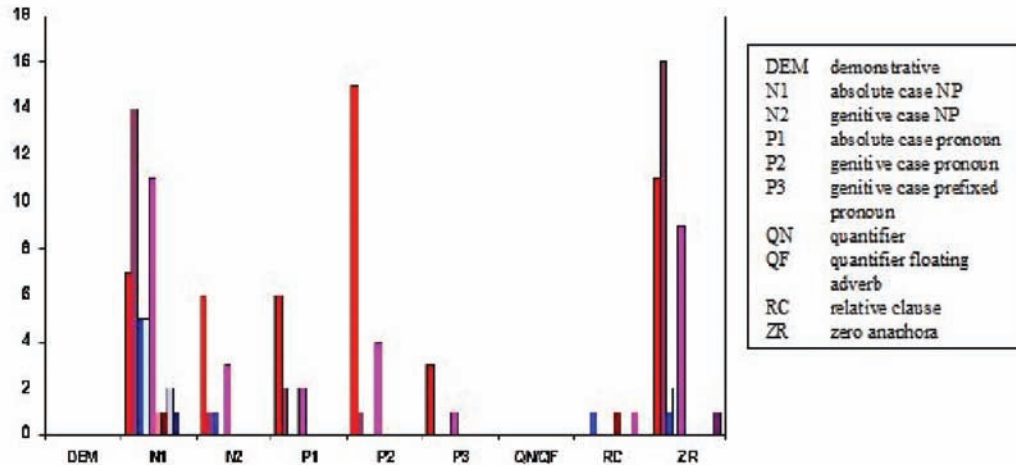


FIGURE 4. NP profile for participants 1-11 in one Pendau text (each color represents a different participant, e.g. red is participant one for each NP type it occurs in)

tives occur, etc. For example, in Pendau typical nouns and pronouns are coded as N1 and P1 respectively for absolute case, and as N2 and P2 for genitive case. Rather than tally these occurrences for each referent they could be tallied for total occurrences in absolute case versus genitive case. This would show in a matrix how many times each participant appears as a pronoun in absolute case, as a pronoun in genitive case, as a typical noun phrase in the absolute case, and as a typical noun phrase in genitive case.

**2.3.2 QUANTIFICATION TOOLS.** The list below summarizes the quantification tools currently available in *Discourse Profiler*. Altogether there are eleven different possible matrices and five sets of matrices of quantified data that can be produced (the sets of matrices are dependent on how many annotations a particular tier has).

Create tally of total NP tokens per participant. This produces a matrix showing how many times each participant occurs for every NP type that has been annotated (e.g. basic NP, pronoun, zero anaphora, etc.)

Create tally of total NP tokens according to the semantic tier tags. This produces a matrix showing how many times each NP participant occurs as a specific category annotated for the semantic tier (e.g. how many times an NP is agent/actor, single argument, patient/undergoer, second object, etc.).

- Create tally of total NP tokens according to the grammatical tier tags. This produces a matrix showing how many times each NP participant occurs as a specific category annotated for the grammatical tier (e.g. how many times a NP is subject, object, left-dislocation, second object, oblique type, etc.).
- Create basic Dooley and Levinsohn S1-S4 statistics for each NP type. This creates a matrix that generalizes the statistics for all the participants. The matrix shows how many occurrences for each of the S1-S4 categories there are for each NP type.
- Create Dooley and Levinsohn S1-S4 statistics for each participant according to all of the possible NP types that have been annotated. This produces a different matrix for each participant. Each participant's matrix shows how many occurrences there are for each of the S1-S4 criteria according to each possible NP type. So if a text has eleven participants, then there will be eleven participant matrices.
- Create Dooley and Levinsohn S1-S4 statistics for each participant regardless of the NP type. This produces one matrix listing all participants and how many occurrences there are for each participant and for each of the S1-S4 criteria.
- Perform basic topic persistence (TP) and referential distance (RD) statistics based on semantic tier tags. This produces one set of matrices for the TP and one set of matrices for the RD for each annotation used in the semantic tier (e.g. actor, undergoer, oblique category, etc.). One set of TP matrices and one set of RD matrices is produced for each semantic annotation. So the number of matrices depends on how many categories have been annotated.
- Perform basic topic persistence (TP) and referential distance (RD) statistics based on grammatical tier tags. This produces one set of matrices for the TP and one set of matrices for the RD for each annotation used in the grammatical tier (e.g. subject, object, oblique, etc.). One set of TP matrices and one set of RD matrices is produced for each grammatical annotation. So the number of matrices depends on how many categories have been annotated.
- Perform advanced topic persistence and referential distance statistics using the Givón method that compares two probable or possible transitive constructions simultaneously. This produces one TP matrix and one RD matrix.
- Perform advanced topic persistence and referential distance statistics using the Dryer method that compares two probable or possible transitive constructions simultaneously. This produces one vertical TP matrix, one horizontal TP matrix, one vertical RD matrix, and one horizontal RD matrix. This method compares the frequency of the actor and undergoer within the same clause according to whether they are the same or lower in frequency.



An additional benefit for the *Discourse Profiler* tags in the *Toolbox* database can be to search for specific examples of text data with a specific code or coding combination. These can also be used for various grammatical analyses and for finding examples to use in research papers or for other research goals (see Pastika 1999, 2006 for examples of an application of this metatagging methodology used in his Balinese texts).

**3. FIELD TYPES.** *Discourse Profiler* includes two field types: information type fields and participant tracking fields. These are described in the following sections.

**3.1 INFORMATION TYPE FIELDS.** It is helpful to record some kinds of linguistic information which can be traced parallel to participant tracking. The idea for the information field originates from Grimes' (1975) description on how to do a 'span analysis' of various discourse information. A span analysis is similar to participant tracking, but is a method of tracing the information flow via spans of information as they appear in clauses. Comparing different spans that appear together, and as they correlate with the participants that are being tracked offers a methodology of identifying patterns where information clusters together. This may help to determine for example where paragraph boundaries occur, or a number of other possibilities.

The information annotated in these fields is typically typological or discourse oriented. Each category of information type must have its own respective field, however the user may have as many different fields as information types that s/he would like to analyze. Typical information types that would be recorded in their own fields are listed below:

- Aspect (e.g. completive versus incompletive; other categories of TAM)
- Clause type (e.g. declarative, interrogative)
- Conversation analysis (this is somewhat limited as the *Discourse Profiler* is designed primarily for use with narratives)<sup>11</sup>
- Dependent clauses (e.g. adverbial clauses, temporal clauses, peripheral elements)
- Direct speech versus nondirect speech (e.g. quoted material)
- Discourse category (e.g. following Grimes (1975), event, nonevent, setting, narrator evaluation)
- Phonetic/phonological features (e.g. loud, soft, aspirated, vowel harmony)
- Propositional relations (connectors; e.g. but, therefore, since)
- Repetition (e.g. tail-head linkage, resumptive repetition, iconic repetition)
- Same subject versus different subject (e.g. Papuan, South American language features)
- Transitivity (e.g. Thompson and Hopper's (1980) categories)
- Unit types (e.g. paragraphs, conversational turns, intonation units (IU) that span more than one clause unit can be marked as in IU1a, IU1b, IU1c, IU2a, IU2b)

---

<sup>11</sup> Turn taking and identification of speakers of a conversation can be 'tracked' through the information type fields (i.e. span analysis) fields. For example, one could trace up to eight speakers by using four fields such as \sp1\_2, \sp3\_4, \sp5\_6, and \sp7\_8. The identification of speakers is then given in the field for a clause, as in '\sp1\_2 speaker2' for a second speaker (or by using the actual name). In the settings within *Discourse Profiler* then for each pair of speakers, only one of the pair of speakers is entered in its own 'list' (i.e. only one item is 'listed').

- Unusual grammatical features (versus typical)
- Verb Spectrum Profile (e.g. Longacre’s approach to discourse analysis (1989))
- Verb types (e.g. specific verb classes in languages which systematically differentiate classes following morphological criteria such as stem formers in Pendau)
- Word Order (e.g. SVO, VOS, SV, VS)

The fields may be commonly abbreviated \it1, \it2, \it3. Alternatively, for information that is widely known to be useful in many languages such as word order, fields can be abbreviated mnemonically: e.g. \wo. The information normally recorded in these fields is typically contrastive binary information such as ‘same subject’ versus ‘different subject’ or information that is normally grouped together into potentially similar or contrasting groups as in various word orders. Typical abbreviations, words, or abbreviated words are entered in these fields. In fact, this information is of a type that is often already entered by analysts. Additionally, specific information found only in an individual language can also be entered in these fields. Multiple fields are used, but the information type for a particular category is restricted to its own respective field.

In summary, there is no closed number to the information type of fields that can be used and the linguist may freely use as many fields as s/he wants to for typological or discourse information as one would normally use within Toolbox. This will most likely be dictated by practical concerns and areas of interest typically worked on in linguistics. This is therefore the more flexible of the two field types. The only constraint is that each category of information must be restricted to its own field. This constraint allows the program *Discourse Profiler* to trace each category of information parallel to the participant tracking in the visual display of a text. This trace or span analysis is useful to locate patterns that are otherwise difficult or impossible to analyze through conventional methods.

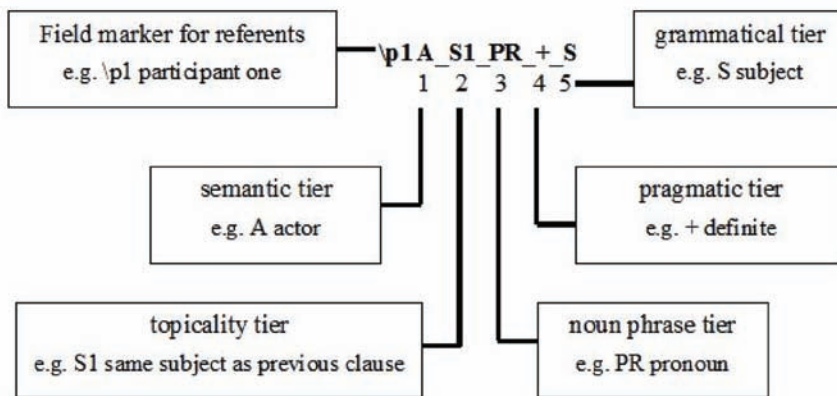


FIGURE 5. The five tier positions used in the participant tracking fields (exact sequence provided by the linguist)

**3.2 PARTICIPANT TRACKING FIELDS AND THEIR FIVE ELEMENTS.** In *Discourse Profiler*, each referent must be uniquely identified with its own field. This is done by using the letter ‘p’ followed by a number: for example, ‘p21’ for participant number 21, and following SIL’s standard format system appears with the backslash marker then as: \p21

There are five elements that need to be coded into one field for each participant (i.e. as requirements of the *Discourse Profiler* software), see Figure 5. These five elements will be referred to as ‘tiers’. These tiers contain autonomous information typically and reflect categories commonly used by the majority of linguists. The information in each of these tiers is abbreviated typically according to the user’s own needs and allows latitude for differing theoretical approaches. The user needs to remember that these tiers are not fields for annotation purposes, and so cannot be left blank. Information type fields can be left blank when there is no relevant information to annotate for a particular clause.

Typically the information encoded for each of the five elements will be an abbreviation (open to the linguist) rather than a word. This is largely because the information within the field needs to be kept down to a practical length (typically one to five characters is enough for each tier). This is primarily for readability reasons, and secondarily for computational reasons. The choice of the abbreviation used for each tier is left up to the linguist (except for use of the topicality tier), although as with many other descriptive methods this needs to be consistent. The important point is to keep the abbreviations for each of these five areas in the same sequence (i.e. 12345, or 54321, etc). Of the five tiers, only the NP type tier is open as to the set of NP types identified. The other four tiers are fairly restricted to what is available to identifying the particular participant as determined by the syntax of each clause. Compare Figure 5 with the following descriptions of each tier.

**3.2.1 SEMANTIC TIER.** For verbal clauses this tier is reserved for the basic macroroles actor and undergoer, or what are often referred to as A and P arguments. Single argument clauses can be further delineated as is commonly practiced in linguistics with the capital S, as the S, A, and P are often the means used to identify differing grammatical systems. This tier is not really meant to be used for semantic roles (e.g. experiencer, instrument, etc.), although with some modification it can work. The reason for this constraint largely has to do with how the calculations of advanced topic continuity statistics are carried out in the *Discourse Profiler* software. For intransitive clauses it may be useful to indicate the difference between undergoer and actor single arguments, as in Su or Sa respectively. For nonverbal clauses this is the tier to use to identify the nominal argument as simply the first or second argument.

**3.2.2 TOPICALITY TIER.** This tier is for a special category of topic continuity in which Dooley and Levinsohn (2001) discuss the importance of whether an NP is activated or not within a particular context. This can be done for the subject (S1-S4) or for the non-subject (N1-N4). The range for subject is as follows (Dooley and Levinsohn 2001: 130):

- **S1** the subject is the same as in the previous clause or sentence
- **S2** the subject was the addressee of a speech reported in the previous sentence (in a closed conversation)

- **S3** the subject was involved in the previous sentence in a non-subject role other than in a closed conversation
- **S4** other changes of subject than those covered by S2 and S3

This is a technique that Dooley and Levinsohn (2001) have developed that comes partly out of treatments on topic continuity (e.g. Chafe 1987, Givón 1983, 1990) and partly from the topic-comment literature (i.e. focus, topic and sentence articulations, e.g. Andrews 1985, Chafe 1976, Givón 1990). The purpose of this technique is to identify the amount of coding material used in each category (especially for S1-S4), and then to determine what coding material is used and the ‘motivations for deviances from default encoding’ (Dooley and Levinsohn 2001: 134). They state for example that (2001: 134),

...common motivations for increased encoding include the presence of a discontinuity and the highlighting of information, while decreased encoding is typically used to identify a VIP.

If this tier category is of no interest then this can be considered to be an optional tier, but due to the constraints currently in *Discourse Profiler* this must be filled in by a dummy character such as with an asterisk. At this stage in the development of the software, other information could be used by the linguist (i.e. annotated here), but there is little point of doing that as there is nothing that *Discourse Profiler* would be able to process.

**3.2.3 REFERENT’S NOUN PHRASE TYPE IDENTIFICATION TIER.** This tier answers the question: what form does the noun phrase take? If it’s a simple noun phrase, then a common abbreviation such as NP may be used. If it’s a basic pronoun, then some other simple abbreviation such as PR may be used. If a case system is used, then another abbreviation can be devised to contrast the noun phrases as such. If a noun phrase is omitted then it is marked as such, for example, if it is due to simple zero anaphora, then ZR may be used to abbreviate it. The choice of abbreviations here can be a rather small set or rather complex depending on the needs of the linguist.

**3.2.4 PRAGMATIC TIER.** This tier answers the question: is the NP definite or indefinite? Alternatively this tier can be used for givenness, specificity, old/new information, etc., however only one category can be used. Definiteness should be the primary consideration here. If the linguist also wants to annotate more than one of these categories, then s/he has the option to use the information type fields and use the tracing feature in *Discourse Profiler*. Technically this is also an optional field, so if one does not want to identify any of these parameters, a dummy character can be used for this tier.

**3.2.5. GRAMMATICAL TIER.** This tier answers the question: what is the grammatical relation (or pivot/non-pivot, etc.) for this participant? This tier typically distinguishes the grammatical subject, object, indirect object, second object, etc. For nonverbal clauses this is the tier reserved for identifying the grammatical function of the noun phrase, typically either as the subject or as the predicated noun phrase. For some languages it may also be useful to annotate the intransitive subject differently from the transitive subject.

These five tiers are coded for each participant of each clause—normally only 2 to 3 referents (and or props) will be coded for each clause. Example (1) provides a typical description of a record with a brief description of what occurs in each field that would typically be used by a linguist for a text. The fields that immediately follow the vernacular text line (\txt) are typically used in interlinearizing a text (e.g. \mr, \ge, \ps). Glosses that occur in the part of speech line (\ps) may sometimes have the same abbreviation used in the NP tier, however this is usually minimal redundancy when and if this occurs. Since the interlinearized portion is usually produced semi-automatically, the requirement to also identify the NP type in the NP tier for each participant is one of little additional time. In the participant tracking fields note that the underline character is used to separate the five different tiers. It is helpful to separate each of the tiers for readability reasons and for computer processing reasons (it will be required once *Discourse Profiler* is released as version 1).

(1)

\ref Text 001	Required record marker
\txt Vernacular text goes here	This is not necessary for Discourse Profiler and is only necessary if the user is working with a Toolbox database
\mr morpheme break line used in Toolbox for example	Used for interlinearizing; optional—not necessary for Discourse Profiler
\ge English gloss	Used for interlinearizing; optional—not necessary for Discourse Profiler
\ps Part of speech	Used for interlinearizing; optional—not necessary for Discourse Profiler
\fte free translation	Optional—not necessary for Discourse Profiler
\wo SVO	Word order of clause—optional in Discourse Profiler but usually for Span Analyses
\it1 Event	Discourse information type 1—contrast for example event and nonevent in the Span Analyses—optional in Discourse Profiler
\it2 SS	Discourse information type 2—contrast for example Same Subject versus Different Subject in the Span Analyses—Optional in Discourse Profiler

<code>\p2 1_2_3_4_5</code>	Necessary for each participant in the clause— numbers represent the five tiers that are used for the abbreviations, e.g. participant 2
<code>\p5 1_2_3_4_5</code>	Necessary for each participant in the clause— numbers represent the five tiers that are used for the abbreviations, e.g. participant 5

Example (2) illustrates a record used for a clause from a Pendau folktale.<sup>12</sup>

```
(2) \rf fktale01.txt 002b
    \pen Ila uo jimo asi mene' negutu
    sanu binaung.
    \mr ila uo jimo asi mene' N-pe-gutu
    sanu binaung
    \ge ABL yonder 3PL/GE just go_up RE-SF/DY-make
    umm lean-to

    \fte After that they just went up to make umm a lean-to.
    \wo SVO
    \it1 Ila uo
    \it3 event
    \p2 A_*_P1+_S
    \p3 A_*_P1+_S
    \p4 A_*_P1+_S
    \p7 P_*_N1_-_O
    \dt 07/Apr/2000
```

In this example the word order is SVO. Information type field one occurs after the word order information type field and is used in this text to identify the discourse connectors (or relators that are used to identify the particular propositional relation between preceding and subsequent information). Next, information type number three identifies the discourse information as an event clause. Information type two is omitted as it is not relevant for this clause. The information type fields are followed by four participants. Participants `\p2`, `\p3`, and `\p4` are three men in this story who have been previously distinguished as distinct referents. Since they are referred to by a plural pronoun *jimo* ‘they’, they are all identified identically with the same five tiers, but identified as distinct by giving them separate numbered fields (see §6.1 for different ways to code plural referents). They are all ‘actors’, and marked with a plus to indicate they are definite. The asterisk indicates that the topicality field is marked with a dummy symbol, i.e. an asterisk in this case. The P1 indicates this is the pronoun used from pronoun set one (or the absolute case). The S indicates that the

<sup>12</sup> Abbreviations used in the Pendau interlinear glossing are as follows: 1SG first singular, 3PL third plural, 3SG third singular, AB absolute case, ABL ablative, COMP completive, DY dynamic verb class, GE genitive case, IV inverse voice, LOC locative, RE realis, RM relative marker, SF stem former, ST stative verb.

referent is the grammatical subject. Referent seven is from noun set one (absolute case), and is indefinite as it is introduced here for the first time in the story. It is an undergoer that functions as the grammatical object.

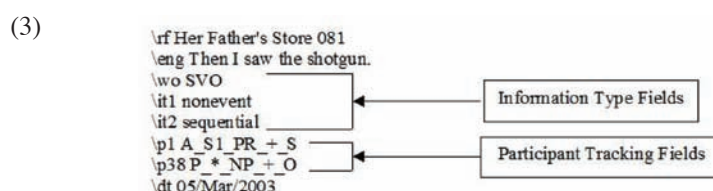
**4. TYPICAL UNIT OF DESCRIPTION.** An important point which must be made is that the typical unit of description is the main clause or what is typically understood to be a ‘simple sentence’. This follows standard practice in descriptive linguistics, and coincides with the choice made in Rhetorical Structure Theory to use the clause as the “elementary discourse unit” (Carlson and Marcu 2001, also see Taboada and Mann 2006). Another important reason for making the basic clause the standard unit is that this works best for participant tracking and follows the methodology practiced for topic continuity analysis (e.g. Givón 1983, 1994). Each database record will typically be a clause, however this does not mean that these units cannot be relative clauses or other subordinate clauses. This is partly a matter of descriptive choice, and partly a matter of practical concerns that are not always easily resolvable. Gildea (1994: 208-211) discusses some of the guidelines needed to determine what is a clause for topic continuity analyses. Difficult decisions often need to be made that often parallel the difficulty of a technical theoretical analysis. On the other hand, because the annotation procedure does not require a rigorous theoretical analysis some room can be made for more practical judgments.

Clausal fragments or interjections often occur in texts and may be incomplete or independent from a syntactic clause. These can still be entered in as if they are a clause unit. If there are participants, these can be tracked through the annotation system. If the interjection is simply a statement by a participant such as ‘yes’, then no participant tracking fields will need to be used. One can still note in the information type fields who the speaker is, if one wants to trace a limited number of speakers. Some adverbial clauses, such as a temporal clause that is dependent on the matrix clause are usually left with the main clause. If one needs to trace adverbial information, then this can also be annotated in a dedicated information type field.

Although the constraint of using syntactic clauses is to be used as the general guideline, the annotation system is flexible enough that it is not mandatory. For example, it is possible to change the units to intonation units or to include sentences such as an if-then sentence that would have two syntactic clauses. This is acceptable to use occasionally, however it is better to indicate the intonation units in the information type fields (or as generally practiced, this can also be indicated in some part of the interlinearized text; see Schultze-Berndt 2006, Himmelmann 2006). If the linguist wants to trace intonation units, this can then be done in the interactive visual map in *Discourse Profiler*. As for sentences with a propositional relation such as an if-then or cause-effect propositional relation, it is better overall for the purposes of the statistical algorithms in *Discourse Profiler* to break these into two clauses, and then trace the semantic components of each clause in an information type field as to its propositional relations. Another related reason for breaking up complex clauses into separate syntactic units is so that the participants can be tracked sequentially or chronologically. If a complex clause has simultaneous actions occurring between different participants, this could perhaps be an exception, and then the participants would all be coded in the same unit. However even an exception like this could potentially skew some of the topic continuity statistics.

The tagging approach is simple as it entails adding only a few extra fields of information to each record in a *Toolbox* database, however it allows for a range from simple to complex ways of coding information concisely. The metatagging system has two types of fields that are important to the linguist: 1) information types (e.g. contrasting event/nonevent, word order, and any number of other grammatical features or discourse information), and 2) participant tracking information for all major referents (this requires 5 elements within each participant/referent field: identification of an NP type, semantic role, grammatical relation, definiteness or specificity, and topicality status).

The example in (3) shows one record from an English fiction story (Quick 1977) with the two types of fields labeled.



The first two fields identify the record, and contain one clause/sentence from this text. Interlinearized text and other annotated fields could occur here as well (see Schultze-Berndt 2006). The information type fields each abbreviate or summarize the particular grammatical or discourse type of information that the linguist wants to document. In this example these are word order (*\wo*), event versus non-event information (*\it1*), and propositional relations (*\it2*). The word order set includes the full range of word orders possible in a language. Although I use a binary set for tracing whether the discourse information in information type one in the example text above is either an event or a non-event (similar to foreground and background), this can be given a more detailed set if one wants to follow for example Hopper-Thompson's set of ten transitivity features (1980) or follow a different approach such as the set from Grimes (1975; Dooley and Levinsohn 2001: 81-83) which includes: event, participant orientation, setting, background (i.e. explanation), evaluation, performative, and collateral. The propositional relations in Pendau for information type two include: cause-effect, concession-contrast, condition-consequence, simultaneous, overlap, alternation, and sequential.

The following fields demonstrate the participant tracking fields. Each referent or participant receives its own unique number, marked in this example as *\p1* and *\p38*. In this sentence participant one is the narrator and is marked with codes which capture these five elements of information: actor/agent (A), same subject as previous clause (S1), pronoun (PR), definite (+) and grammatical subject (S). The shotgun referent's code can be interpreted as: patient/undergoer (P), unmarked for topicality (\*; or not applicable), a typical noun phrase (NP), definite (+) and the grammatical object (O).

Although this metatagging system was developed for the *Discourse Profiler* software, it is not necessary to use this software in order to take advantage of it for annotating or archiving texts. The metatagging system also takes advantage of the capabilities of *Toolbox*, however it is not necessary to use *Toolbox*. The choice remaining would be to use a word processor (or basic text editor) that can save the data in plain text format. This extraordi-



nary choice would still necessitate using the field markers in the database format developed by SIL, and as used in *Toolbox*. Although this is not the preferable way to work, there may be some circumstances that preclude using *Toolbox*. This approach then still allows the linguist or language worker to annotate texts for documenting a language for archival purposes, and allow the possibility of further analysis of a tagged text using *Discourse Profiler*. There may well be another software available in the future or one that could be adapted to using the kind of database structure used in *Toolbox*, and this as well would likely be for exceptional reasons.

**5. TOOLBOX AND DISCOURSE PROFILER.** There are three features in *Toolbox* that especially are of help in annotating a text or working with the finished annotated text. These features are the semantic range feature, the browse mode feature, and the filtering feature.

**5.1 SEMANTIC RANGE.** This feature is helpful for staying consistent for delimiting the range of abbreviations or words used in a particular field while one is entering the annotation. It works similarly to a spell checker, and not only allows the words or abbreviations already kept in a special list for that particular field, but allows new additions easily as needed. This feature can be particularly useful for the participant tracking fields, as there may easily be twenty or more possibilities coded for a participant. Once an initial list of possibilities is listed for a particular participant, the semantic range feature provides a running list that can be used as a menu to choose from. This means there is less typing to perform.

**5.2 BROWSE.** This mode is helpful for doing some basic discourse analysis, in addition to typical syntactic analysis often done for descriptive purposes (also useful when looking for examples to use in a paper). For example, in the browse mode the user can view multiple records with selected fields displayed in columns. This feature allows one to do some basic span analysis of some information types such as word order. It can also be used to do some limited participant tracking, but for texts with a large number of referents it is not practical.

**5.3 FILTER.** This feature is one of the most powerful features of *Toolbox*. This feature allows the user to delimit a particular database to include and/or exclude the exact information desired to be viewed. For example, if one wants to view only the records of a text which have an SVO word order, then once this is specified in the filter then only those clauses can be viewed and studied. When the filter is turned off, then the entire database is once again viewable. Many other types of filters can be set up that range from simple to complex (including combining information in multiple fields that should be viewed or not viewed). The filtering capabilities are enhanced even more with the large number of choices that the metatagging method introduced here allows. Finally combining the browse mode and the filtering capabilities allows for even greater research capabilities.

Semantic Tier		NP Identification Tier		Grammatical Tier		Topicality Tier		Pragmatic Tier	
A	agent (actor)	PR	pronoun	S	subject	\$1-\$4	codes used for Levinsohn & Dooley analysis	+	definite
P	patient (undergoer)	NP	noun phrase	O	object	*	dummy symbol	-	indefinite
LOC	locative oblique	P1	pronoun set 1 (absolute case)	Q	oblique				
ABL	ablative oblique	N1	noun phrase set 1 (absolute case)	O2	second object				
O2	second object (undergoer)	N2	noun phrase set 2 (genitive case)	I	intransitive subject				
EQ1	first equative noun phrase	N1/RC	noun phrase set 1 as head of relative clause	Pred	predicate of a verbless clause				
EQ2	second equative noun phrase	RC	relative clause	Qpred	oblique predicate of a verbless clause				
Narg	non-predicate noun phrase argument for verbless clause	ZR	zero anaphora	Left-d	left-dislocated noun phrase				
EX1	first existential clause's noun phrase	P2	pronoun set 2 (genitive case)	APPOS	appositional noun phrase				
EX2	second existential clause's noun phrase	P3	pronoun set 3 (inverse prefix, genitive case)						
Exarg	existential clause's argument	P1/RC	pronoun set 1 as head of relative clause						
C	complement clause functioning as a clausal argument								
APPOS	appositional noun phrase								

FIGURE 6. Examples of abbreviations used for annotations in Pendau participant tracking

**6. EXAMPLES FROM PENDAU.** This section lists several sample records according to various grammatical categories drawn from the endangered Pendau language (see Quick 2003, 2007). These are representative examples of various kinds of clauses or other categories. They are only illustrative here and not necessarily definitive in how the metatagging approach may be used to annotate texts. Also note that typically there are separate interlinearized lines for part of speech (*\ps*) and the gloss (*\ge*). I have merged these two lines into the gloss line (*\ge*) to simplify the examples for the presentation. Figure 6 lists the abbreviations used in the Pendau participant tracking fields.

**6.1 PLURAL NPs.** Example (4) illustrates two different cases for handling plural referents. The plural pronoun is used for two referents in this story about the monkey and the turtle. These are coded in separate fields as participant one and participant two, but have the exact same annotation. They are both actors and the grammatical subject of the clause. They are also marked as definite as they have already been introduced prior to this clause (as is typical of pronouns). The third referent 'fish' could be singular or plural just as it is in English. The approach I generally take is not to indicate plural referents as multiple referents in the annotation unless they are unpacked later in the text as distinct referents. So in

this text the ‘fish’ searched for in this first instance is indicated as an indefinite noun phrase and is the undergoer and grammatical object of the clause. This is a rather common feature of minor ‘props’ in a text, and there may also be later instances of ‘fish’ being searched for which are probably not the same fish. The fourth referent annotated in this clause is the oblique introducing the river.

```
(4) \rf turtle.pin 003b
    \pen Jimo      ma'o nelolo          bau ribangkalang.
    \mr jimo      ma'o N-pe-lolo       bau ri=bangkalang
    \ge 3PL/AB    go RE-SF/DY-search_for fish LOC=river

    \fte They went to search for fish in the river.
    \wo SVOQ
    \p1 A_*_P1+_S
    \p2 A_*_P1+_S
    \p3 P__N1_-_O
    \p4 LOC_*_N1_-_Q
    \dt 24/Apr/2000
```

**6.2 INTRANSITIVE CLAUSES.** Example (5) illustrates the second half of a coordinate sentence that is a stative clause. The single argument is an undergoer. Grammatical subjects are marked with a capital I (for Intransitive grammatical subject). This is an important distinction in the annotation system in order to distinguish intransitive clauses from transitive clauses in the tabulation of topic continuity statistics. In Pendau the stative subject is marked with the absolute case (i.e. common nouns are unmarked in the absolute case).

```
(5) \rf Daras_fish_story05.035d
    \pen o barumbang noogemo.
    \mr o barumbang no-oge=mo
    \ge and wave ST/RE-large=COMP

    \fte ...and the waves were huge.
    \wo SV
    \it1 event
    \p28 P_*_N1+_I
    \dt 11/Oct/2006
```

**6.3 VERBLESS CLAUSES.** Example (6) illustrates an equative clause. As with other verbless clauses one argument can be analyzed as the grammatical subject, and the other noun phrase as the predicate. One approach to annotating the word order is simply to designate one noun phrase as number one (e.g. EQ1) and the second noun phrase as number two (e.g. EQ2). Since an equative clause is typically a description it is noted here as nonevent information. The decision made for annotating the word order can now also be used for the participant tracking field. Since the equative clause refers to the same participant it is useful to differentiate in the semantic tier between the two noun phrases. Although one could identify verbless clauses as an ‘intransitive clause’ (e.g. see Dixon 1988: 63-68), for the metatagging system presented here it is not necessary to identify a semantic role (which in

any case would be simply a ‘single argument’ (or ‘S’ as contrasted with ‘A’ and ‘P’). It also may be more often helpful in the various analytical methods available in *Discourse Profiler* to maintain a separation for the coding between verbal clauses and verbless clauses. For the grammatical tier there are probably several possibilities how these can be indicated. In this example I have marked the first equative noun phrase (EQ1) as the subject (S), and the second equative noun phrase (EQ2) as the predicate (Pred).

```
(6)  \rf Daras_fish_story05.028
      \pen Bau      tono'uore          uo,      topenyo      repa.
      \mr bau      to=no'u-ore        'uo      tope=nyo     repa
      \ge fish      RM=1SG.IV/RE-pull  yonder   name=3SG/GE  snapper

      \fte The fish that I pulled up there, it's name is a snapper.
      \wo EQ1_EQ2
      \it1 nonevent
      \p9 EQ1_*_N1+_S
      \p9 EQ2_*_N1_-_Pred
      \dt 06/Oct/2006
```

**6.4 FLEXIBILITY.** The *Discourse Profiler* metatagging system can be very flexible to meet individual needs. For example, Pastika (1999, 2006) intended to use this metatagging system to work on topic continuity analyses of Balinese texts, and instead of using five elements of information in the participant tracking fields, he used four elements. However, since he still closely followed the constraints of the metatagging system, it is still possible to take his data and simply add a dummy character at some consistent point to make up for the deficiency (since *Discourse Profiler* currently requires five elements in each participant field). He also shows another possibility for adapting the metatagging system in demonstrating in his database that one does not need to label the annotation fields as required by the software, and even this can be deleted and manipulated with a global automatic change in order to make his database work in the current version of *Discourse Profiler*. This also highlights the possibility of changing *Discourse Profiler* in a future version (in its second generation perhaps, or sooner) in order to make the program more flexible to allow optional elements in these fields.

**7. DISCOURSE PROFILER IN THE FUTURE.** The *Discourse Profiler*'s metatagging system presented here is offered as a new tool to add to the fast growing inventory of ways to annotate endangered languages' texts. The first advantage is that it leverages the software tool *Toolbox* which many linguists are already using (and for which there are many people who can help new users). The second advantage is that the metatagging system presented here is a multipurpose system. Although the metatagging system was developed along with the development of *Discourse Profiler*, the separation of the annotated text database in *Toolbox* that is stored as a plain text from the proprietary *Discourse Profiler* software fulfills the current criteria for archiving texts.

The software developed for *Discourse Profiler* is still in its first generation, and the metatagging system has not been exhaustively tested. A current drawback of the *Discourse Profiler*'s metatagging system is that it is still necessary to enter the metatagged data manu-

ally into *Toolbox*. As has already been mentioned, there is some help from *Toolbox* if one uses the semantic range fields, however this approach is still not the ideal way to enter an annotation. The semantic range fields can be used for both types of fields when needed or helpful.

Since the span analysis fields are highly abstract, there does not seem to be a lot of potential for developing an automated tagging feature useful for all languages. There may be individual information type fields that can be created through the use of macros for certain types of information that could theoretically be drawn from an interlinearized text, but this will likely have to be left up to individual situations.

In the near future I expect to develop an automated tagging feature for the participant tracking fields. The reason there is strong potential for this is based on the fact that texts are structured and that the five tiers of information provide the basis for a rich variety of analyses as demonstrated by the *Discourse Profiler* software. These analyses can theoretically be reversed so to speak, and used as algorithms to build a kind of weak artificial intelligence that allows the tagger to make 'guesses' of the most likely annotations that would be needed for a given participant in a text. Essentially the tagger would provide a short list of the most likely tags for the participants of a clause.

As more languages are documented, the more likely it will be that some of the texts will be annotated with discourse information when there is an easy to learn system that provides a solution for complementing the current archiving goals for endangered languages' texts. I propose that the metatagging system presented here as developed for the *Discourse Profiler* can serve as a current robust solution to the current gap in annotating texts with discourse information. The potential for an automated tagger would of course dramatically increase the capability for paving the way for annotating texts with discourse information more rapidly, allowing for the possibility of increasing the potential number of texts archived for any particular language with these annotations.

The ultimate goal that has been focused on in this paper has been to offer another tool for documenting languages with an additional breadth of information that can be used for the conservation and preservation of endangered languages. There is also the bonus that this system provides the added potential of performing discourse analysis of texts as well as a variety of other linguistic analyses which should make it more attractive to an even wider range of linguists.

Adding the discourse-text annotations can be kept to a minimum load of additional work, and I am proposing that at least some texts of a corpus that will be archived should be annotated with some discourse-text information. Since one of the goals of archiving texts from an endangered language is to provide a robust documentation of a language, I propose that this software and metatagging system provides a method that can contribute to this important task and that it complements current proposals (e.g. Gippert, Himmelmann and Mosel (eds.) 2006).

## REFERENCES

- ANDREWS, AVERY. 1985. The major functions of the noun phrase. In *Language typology and syntactic description, Clause structure*, Shopen, ed. , Vol. I. Cambridge: Cambridge University Press.
- Carlson, Lynn and DANIEL MARCU. 2001. Discourse tagging reference manual. Unpublished manuscript. <http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- CHAFE, WALLACE L. 1976. Givenness, contrastiveness, definiteness, subject, topics, and point of view. In *Subject and Topic*, Charles N. Li, ed., pp. 25-56. New York: Academic Press.
- CHAFE, WALLACE L. 1987. Cognitive constraints on information flow. In *Coherence and Grounding in Discourse*, Russell S. Tomlin, ed., pp. 21-51. Amsterdam: John Benjamins.
- DIXON, R. M. W. 1988. *A grammar of Boumaa Fijiaan*. Chicago: University of Chicago Press.
- DOOLEY, ROBERT A., and Stephen H. Levinsohn. 2001. *Analyzing discourse*. Dallas: SIL International.
- DRYER, MATTHEW. 1994. The discourse function of the Kutenai inverse. In *Voice and Inversion*. Talmy Givón, ed. Amsterdam: John Benjamins Publishing Company.
- EDMONDSON, JEROLD A. and DONALD A. BURQUEST. 1992. [1998, third edition] *A survey of linguistic theories*. Dallas: The Summer Institute of Linguistics.
- GIPPERT, JOST, NIKOLAUS P. HIMMELMANN and ULRIKE MOSEL, eds. 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- GILDEA, SPIKE. 1994. Semantic and pragmatic inverse: “Inverse Alignment” and “Inverse Voice” in Carib of Surinam. In *Voice and Inversion*, Talmy Givón, ed. Amsterdam: John Benjamins Publishing Company.
- GIVÓN, TALMY. 1994. The pragmatics of de-transitive voice: Functional and typological aspects of inversion. In *Voice and Inversion*, Talmy Givón, ed. Amsterdam: John Benjamins Publishing Company.
- GIVÓN, TALMY. 1990. *Syntax a functional-typological introduction. Vol. II*. Amsterdam: John Benjamins.
- GIVÓN, TALMY, ed. 1983. *Topic Continuity in Discourse: A Quantitative Cross-language Study*. Amsterdam: John Benjamins Publishing Company.
- GIVÓN, TALMY, ed. 1994. *Voice and Inversion*. Amsterdam: John Benjamins Publishing Company.
- GRIMES, JOSEPH E. 1975. *The Thread of Discourse*. The Hague: Mouton.
- HIMMELMANN, NIKOLAUS P. 2006. The challenges of segmenting spoken language. In *Essentials of Language Documentation*, Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, eds. Berlin: Mouton de Gruyter.
- HOPPER, PAUL J. and S. A. THOMPSON. 1980. Transitivity in grammar and discourse. *Language* 56.251-99.
- LONGACRE, ROBERT E. 1983. *The grammar of discourse*. New York: Plenum Press.
- LONGACRE, ROBERT E. 1989. Two hypotheses regarding text generation and analysis. *Discourse Processes* 12: 413-460.

- LONGACRE, ROBERT E. 1996. *The grammar of discourse*. Second edition. New York: Plenum Press.
- PASTIKA, I WAYAN. 1999. Voice selection in Balinese narrative discourse. PhD diss., Australian National University.
- PASTIKA, I WAYAN. 2006. Voice selection in Balinese narrative discourse. Denpasar, Bali: Pustaka Larasan.
- QUICK, PHIL. 1977. Her father's store. Unpublished fiction short story.
- QUICK, PHIL. 1996. Multilinear discourse analysis software demonstration. In H. Andrew Black, Alan Buseman, David Payne, Gary F. Simons (eds.), *Proceedings of the 1996 General CARLA Conference, November 14-15, 1996*, pp. 291-309. Waxhaw, NC/Dallas: JAARS and Summer Institute of Linguistics.
- QUICK, PHIL. 1997. Active and inverse voice selection criteria in Pendau, a Western Austronesian language. In *Proceedings of the Seventh International Conference on Austronesian Linguistics*, pp. 461-482. Amsterdam/Atlanta: Editions Rodopi B.V.
- QUICK, PHIL. 2002. A sketch of the primary transitive verbs in Pendau. In *The historical and typological development of Western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross. Canberra: Pacific Linguistics.
- QUICK, PHIL. 2003. A grammar of the Pendau language. PhD diss., Australian National University.
- QUICK, PHIL. 2005. Topic continuity, voice, and word order in Pendau. In *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, I Wayan Arka and Malcolm Ross, eds., pp. 221-242. Canberra: Pacific Linguistics.
- QUICK, PHIL. 2007. A grammar of the Pendau language of Central Sulawesi, Indonesia. Canberra: Pacific Linguistics.
- SCHULTZE-BERNDT, EVA. 2006. Linguistic annotation. In *Essentials of Language Documentation*, Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, eds. Berlin: Mouton de Gruyter.
- TABOADA, MAITE and WILLIAM C. MANN. 2006. Rhetorical Structure Theory: Looking back and moving ahead. Unpublished manuscript. [http://www.sfu.ca/rst/pdfs/Taboada\\_Mann\\_RST\\_Part1.pdf](http://www.sfu.ca/rst/pdfs/Taboada_Mann_RST_Part1.pdf)

Phil Quick

[phil.quick@sil.org](mailto:phil.quick@sil.org)







Language Documentation & Conservation Special Publication No. 1

ISBN 978-0-8248-3309-1