

# A Conceptual Framework for Developing e-Learning Materials from Endangered Language Documentation<sup>1</sup>

**Meng-Chien Yang**

Department of Computer and  
Communication Engineering  
Providence University, Taiwan

mcyang2@pu.edu.tw

**D. Victoria Rau**

Department of English Language,  
Literature and Linguistics  
Providence University, Taiwan

dhrau@pu.edu.tw

To appear in *Language Resources and Evaluation special issue on Asian language technology*, Eds.  
by Tokunaga Takenobu and Chu-Ren Huang

## Abstract

Preservation of an endangered language is an important and difficult task. The preservation project is proposed to include documentation, archiving and development of shared resources for the endangered language, with an underlying goal of language revitalization among the younger generation. In this paper, we propose a framework that connects three different tasks related to language documentation: language archiving, language processing and creating e-learning materials. We are using this framework to document and produce e-learning materials for the Taiwanese aboriginal language Yami. The proposed framework has the potential to become an effective tool for documenting and developing e-learning programs for other endangered languages in Asia.

## 1 Introduction

The impact of globalization and urbanization has caused the extinction of many aboriginal languages on our planet. This language death process not only reduces the number of native languages but also wipes out the cultural heritage connected with those languages (Xu 2001). Therefore, preservation and archiving of these endangered native languages is both vital and critical. Many projects around the world are seeking to preserve endangered native languages (e.g., Lublinskaya 2002; Psutka 2002).

The attempt to preserve an endangered language includes several steps: documenting and recording the oral and written literature, compiling the grammar and a dictionary of the language, and annotating the documentation related to this language. It is also important to find an effective approach to teach the endangered language to the ethnic group using the language, particularly to members of the younger generation. Although in East Asian countries these young indigenous people often live in urban areas without any connection to their place of origin, many have easy access to internet technology. This could be used to provide a link to their ancestral language and culture, but too often the internet, like the culture around them, provides access only to the dominant language.

According to a study by Whaley (2003), the important factors required to help an endangered language survive include:

1. a well developed preservation and maintenance program for the language;

---

<sup>1</sup> The preliminary version of this paper was presented at the ALR workshop, Jeju Island, Korea (Yang & Rau 2005). We are also thankful to the helpful suggestions and comments from three anonymous reviewers.

2. use of information technology in the preserving project;
3. a new world order, especially economic and political shifts;
4. an environment for learning and exploring the language.

Thus it is important that an endangered language preservation and documentation project be carefully and comprehensively planned. This project needs to take advantage of state-of-the-art technologies and establish an environment conducive to learning. The goal of our project is not only to document and archive the endangered language, but also to develop an environment that fosters learning of the language.

The endangered language under study is Yami, a Philippine Batanic language, spoken by less than 4000 people. As a Philippine language, Yami displays the typical Philippine “focus” system. The verb form changes to reflect the semantic role of the “Subject” of the sentence: Agent, Patient, Location, and Instrument/Benefactive. Yami affixation manifests the three features characterizing Agglutinating Languages: (1) a word consists of a root and several affixes, (2) the root and affixes are relatively easily separated, and (3) each affix generally has only one meaning (Rau & Dong 2006). According to Rau’s (1995) sociolinguistic survey on Orchid Island, Iraralay is the only community of the six villages on the island where children still use some Yami for daily interaction. Although Yami has been offered as an elective in elementary school since 1998, Yami is gradually being replaced by Mandarin Chinese.

In this paper, an integrated framework for building an e-learning environment by using the digital archiving of the endangered language is proposed. Our targeted learners at the initial stage of the development are Chinese university students learning Yami as a second language in Taiwan (Rau, Yang & Dong 2006) with the ultimate goal of reaching Yami youngsters who do not live in their language community. Although the complete work of documentation will take many years, the Yami language is in danger of being lost due to contact with Mandarin Chinese and migration out of the community. Therefore, we have developed a strategy to make language items available in learning materials as soon as they have been collected, taking advantage of information technology and computer networking. Using these technologies we have developed an integrated platform for documenting, processing and learning that will help both Yami youngsters and other students taking Yami as a second language.

The collection of the Yami language data began in 1994. The second author collaborated with a Yami consultant to develop an extensive corpus, which has been the basis of extensive research on the language. The framework proposed in this paper is being used in an ongoing grant-supported project for archiving and documenting the Yami language (ELDP, MDP0114). The first materials archived included 20 narratives, a reference grammar, and trilingual dictionary with 2000 entries (Rau & Dong, 2006), and multimedia pedagogical materials (Rau et al. 2005). We are now collaborating with local consultants to document daily conversations, business transactions, festivals, and ceremonies. The topics were selected based on consultation of previous research on Yami ethnography, and are designed to meet the standards stipulated by the R.O.C. Ministry of Education for developing Austronesian teaching materials in Taiwan. The topics are closely related to those selected for inclusion in four volumes of Yami multimedia teaching materials the second author is currently developing. The Yami digital archiving and the Yami e-learning conducted under the proposed framework are available at the following web sites: <http://yamiproject.cs.pu.edu.tw/yami/> and <http://yamiproject.cs.edu.tw/elearn/>.

Currently there are nearly 30 ongoing endangered language documentation projects in the Asia and Pacific Areas. These projects are mostly supported by governmental researching funding or international grant organizations, such as ELDP and DoBes. The first digitalizing project on Formosan languages, i.e., the Formosan Language Archive at Academia Sinica, has produced the first online Formosan corpora (Zeitoun et. al. 2003, Zeitoun & Yu 2005). Most projects aim to create shareable endangered language resources. Many studies have provided valuable guidance for language documentation (Bird & Simons 2003a, Bird & Simons 2003b) and development of multimedia learning materials (Csató & Nathan 2002). However, to our knowledge, no study has attempted to provide an

integrated conceptual framework for the whole process from documentation to e-learning of endangered languages. It is this gap that the current research aims to fill. In this paper, we describe the collaboration between linguists, computer scientists and the ethnic community to create a framework for developing e-learning materials based on endangered language documentation. We believe this proposed framework will serve as an important guideline for future projects on digital archives of endangered languages.

This paper addresses the needs of a field linguist who has collected some endangered language materials from an indigenous community and intends to establish a shareable language resource via the internet. In the following discussion, the steps are provided for transforming the linguistic framework of the documentation project into e-learning materials. Section 2 illustrates the proposed framework and gives explanations of our design. Section 3 discusses the roles of the three different groups of participants collaborating in the endangered language documentation project. Section 4 is the conclusion.

## **2 An Integrated Framework for Developing E-learning Materials of Endangered Languages**

In this section, we describe the proposed framework and the connections between endangered language documentation and e-learning development. The endangered language documentation project is considered an important step for language revitalization. The purpose of documentary linguistics, as stated in Himmelmann (1998), is to collect comprehensive records of the endangered languages. These records can support different linguistic research and can be used as comprehensive descriptions of the endangered languages. With the help of computer-related technologies, linguists can use digitization and computerized tools to preserve and manipulate this documentation. The documentation project can produce a detailed and well-documented database and even web pages for endangered languages. Many endangered language documentation projects have created a comprehensive language database which can be accessed via internet. Revitalization of endangered languages is more difficult to achieve, however, and faces many challenges. In particular, it is very expensive to create learning materials for the endangered languages in the dominant language environment. For example, the Formosan Language Archive at Academia Sinica has not reached its goal of developing a pedagogical tool due to lack of funding (Zeitoun and Yu 2005: 168). In comparison with print materials, computer mediated technologies such as multimedia and the use of interactive web technology can create a relatively inexpensive yet advanced tool for learning of endangered languages (Eisenlohr 2004).

Successful development of multimedia learning materials for endangered languages requires many different domains of knowledge: linguistics, information technology (IT) and indigenous knowledge of the endangered language community. Many studies have indicated the importance of community involvement in the development of multimedia learning materials for the endangered languages. Csató & Nathan (2003) showed the participation of the local language community in developing a multimedia CD-ROM for the Karaim of Lithuania produced high quality learning materials. Similarly, the assistance of Yami consultants on the Yami documentation project has been vital to production of complete and useful Yami language e-learning materials.

The IT professionals who participate in the endangered language documentation and e-learning development project not only give assistance and technological support to the documentary linguists but must collaborate with other team members closely. Our experience showed that the knowledge of the IT professionals convinced the Yami community members that e-learning and multimedia tech-

nology could help Yami learners achieve competence<sup>2</sup>. As a result, the Yami community members started to show interest in developing e-learning materials for their language. The design of our e-learning program for the endangered language is tied to the results of the language documentation project and integrates related multimedia content using IT technology. For example, several dialogues of the language lessons and the learning activities have been constructed based on the documented narratives. The cross-disciplinary team members have worked collaboratively to develop e-learning programs for the endangered language.

Incorporating all these elements, an integrated framework for developing an e-learning program for an endangered language is illustrated in Figure 1. In this framework, language documentation and e-learning programs are two separate modules. The language documentation module has two phases: preparation and digitalization. The e-learning programs 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.

The proposed framework focuses on describing the relationships between language documentation and e-learning program development. As we believe an e-learning program for the endangered language will be a useful tool for revitalizing the endangered language, it is important to plan and prepare the e-learning programs at the earliest step of the language documentation project. Collaboration of the three groups of participants will yield the best e-learning program.

The following subsections describe each module of the framework.

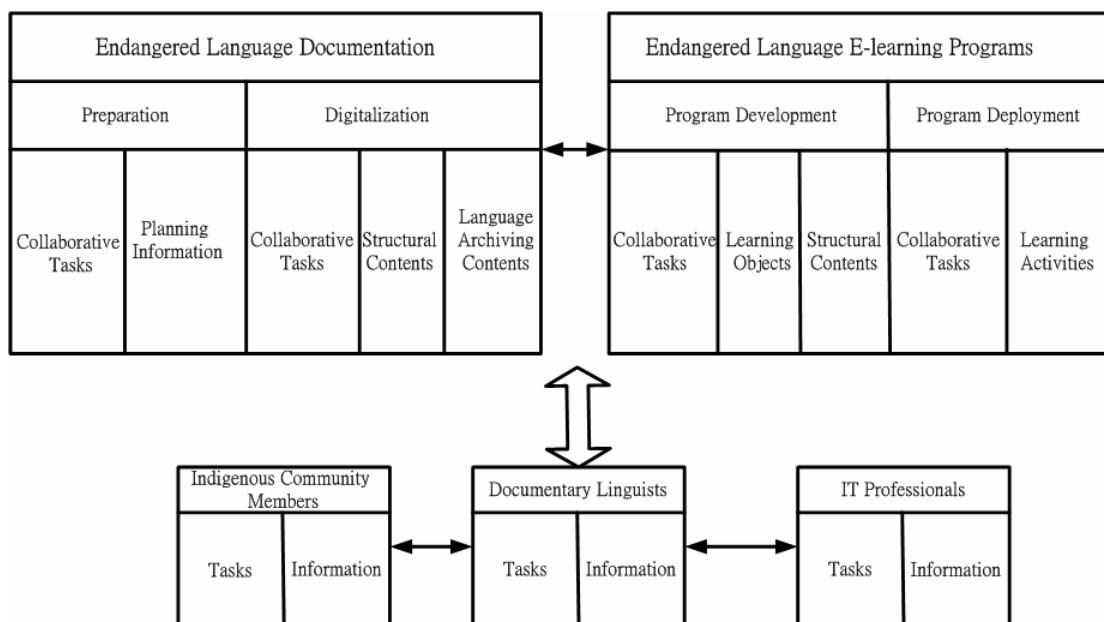


Figure 1 Framework for Endangered Language Documentation and e-Learning Development

## 2.1 Endangered Language Documentation Module

<sup>2</sup> We led a workshop to promote Yami e-learning in April 2006 on Orchid Island and received positive feedback from the local participants.

If the documentary linguist in an endangered language documentation project accepts e-learning as an objective of his/her project, how does s/he proceed? Clearly, digital archiving of language documentation is different from traditional documentation projects. The multimedia digital archiving includes both texts and the multimedia data related to the corpus. For the multimedia data, all formats, including recordings, video clips, images and animations, are considered. Since the amount of multimedia data is very large, how does one decide which is to be used in developing the e-learning program for the endangered language? Furthermore, considering that developing multimedia learning materials is very expensive and time-consuming (Nathan 2004), how can one find sufficient funding for e-learning? After the corpus of an endangered language is built, how are the e-learning programs created from the language corpus? How does the linguist collect more suitable corpora for improving the multimedia learning materials?

The following paragraphs attempt to address the above-mentioned questions. We begin by discussing the two phases in language documentation.

### **2.1.1 Preparation Phase**

The preparation phase involves organizing a team for the language documentation and finding feasible methods to develop an e-learning program. Designing the multimedia learning materials for the endangered language is probably a new idea for all the participants involved. Hence, the collaborative tasks at this stage may include the following:

- Collecting information about the internet infrastructure in the indigenous community: Many regions in Asia have access to the internet in different bandwidths. The team should work together to find out whether or to what extent internet is available in the indigenous community.
- Developing a prototype or preliminary presentation of the e-learning programs: The team works together to develop a visual sketch of the e-learning program.
- Listing available resources: The team searches for possible resources and funding to support the development of the e-learning programs.
- Delivering the e-learning development plan: The team works together to decide on a draft plan and timeline for developing the e-learning programs.

In summary, the preparation phase aims to mobilize a project team. The project leader should adapt the list of the collaborative tasks and information needed to suit the particular context.

### **2.1.2 Language documentation phase**

The language documentation phase is the core of the endangered language project. Many studies have described the practices, software tools, metadata and digital archiving involved (e.g., Bird & Simons 2003a, Bird & Simons 2003b, Himmelmann 1998). Here, we focus on describing those tasks and contents that are helpful in developing e-learning programs for the endangered languages. As in the preparation phase, there are several important tasks of the language documentation phase that should be jointly carried out by the three groups of participants:

- Defining an approach for acquiring the multimedia documentation: if development of e-learning programs is one of the major goals of the project, following Johnson's suggestions (2003), it is important to develop strategies for quickly accumulating multimedia repositories of the language from the field work.
- Reviewing and validating the multimedia data: it is critical that the contents of the collected multimedia data are checked and annotated by proper authorities in the team, i.e. the linguists and community members. Finally, it is important that the digital quality of the multimedia data be reviewed by the IT professionals.
- Searching for an effective e-learning method: although there are numerous e-learning approaches for learning socially dominant languages, there are very few studies on learning endangered languages (Csató & Nathan 2003, Nathan 2004, Rau & Yang 2006, Ward & van Genabith 2003). An effective way is for the IT professionals to analyze various learning approaches and work with the linguists and the indigenous members to find a suitable and effec-

tive e-learning approach for the endangered language. Later, the e-learning instructors and the learners should give feedback to adjust the approach to incorporate best educational practices. Our study illustrated the steps for this practice (Rau & Yang 2006).

The fundamental structural contents of the e-learning program include: (1) Metadata: essential for describing and sharing the digital language repositories, (2) Grammar structure and elicitation structure: essential for describing the notation and textual representation of the language, (3) Multimedia database schema: essential for describing the multimedia language data. Although the ideal data is a complete corpus of the endangered language, a more obtainable goal is to select higher priority items to be documented first. The selection of these items can be integrated with the task of developing an effective learning method.

## **2.2 E-learning program for endangered language module**

The e-learning program in the framework is a web-based multimedia learning program for the endangered language. If the indigenous community does not have access to internet connection, the content of the e-learning program can be reproduced and delivered as a multimedia CD-ROM. Because of the critical status of endangered languages, development of e-learning programs should start as soon as a small set of multimedia language data has been collected and documented. In the following paragraphs, we discuss the two phases of the e-learning program.

### **2.2.1 Program Development Phase**

The contents of the e-learning programs originate from the corpus of the documentation project. The collaborative tasks of the program development phase are as follows:

- Designing the presentation and user interface of the e-learning program: the presentation of the e-learning program is crucial to attract potential learners to use this system. The design of the user interface determines the efficacy of the program. The IT professionals are responsible for presenting multiple designs and making necessary adjustment after receiving feedback from the other two groups of participants.
- Creating and organizing learning objects: An e-learning program contains several learning objects or modules that can be reused to achieve a particular learning purpose. A learning object contains descriptive web pages to describe the learning materials, evaluative web pages for self-testing the learning results and multimedia description web pages to describe language terms used in the images, videos or animations. The IT professionals can create templates for developing the learning objects for the endangered language.

The e-learning program and the supportive database to store these e-learning programs can be stored and managed using commercial e-learning platforms. However, at the beginning stage of developing the e-learning programs for an endangered language, we recommend using inexpensive freeware servers, such as the Apache. In addition, an online version of the dictionary and grammatical explanation of the endangered language are two essential tools for e-learning programs and should be developed concurrently.

### **2.2.2 Program Deployment Phase**

When the e-learning programs are ready to be launched, the next task is to promote the use of the e-learning programs while continuously evaluating the effectiveness of the learning activities. The e-learning programs can not be considered successful unless the number of users is substantial and the learning experiences of the users are satisfactory. Thus the targeted group in the initial stage of deployment is the mainstream community in Taiwan, as illustrated in Rau, Yang, and Dong (2006). However, plans are being made to conduct needs analysis of the youngsters on Orchid Island to de-

velop e-learning program to meet their goals. A follow-up project at Providence University is currently being conducted to evaluate the effectiveness of the e-learning program for Chinese learners.

The collaborative tasks in this phase are summarized as follows:

- Conducting e-learning workshops in the indigenous community<sup>3</sup>,
- Creating promotional web sites to attract more people to use the e-learning program,
- Conducting evaluation of the effectiveness of the program.

In summary, the program development phase shows the conceptual foundation for developing the multimedia e-learning program. During the program deployment phase, the e-program is used as a tool for learning and revitalizing the endangered language.

### 3 The Roles of Different Participants

In this section, we discuss the roles of the three groups of participants in the framework: (1) linguists who are familiar with language documentation and linguistic description of the endangered language, (2) indigenous community consultants who have communicative competence in the endangered language, and (3) the IT professionals who can help transform the results of the language documentation into an effective e-learning program.

In the proposed framework, a minimum of one linguist, one language community member and one IT professional are needed to form the project management team. As shown in Figure 1, each group of participants is assigned a list of tasks. For a well-funded project, we would suggest that the team also include sociolinguists, language teachers, web designers, and multimedia graphic designers. The following subsections describe the roles and the characteristics of the three groups of participants.

#### 3.1 The Role of the Documentary Linguist

The documentary linguist is the leader of the language documentation project. S/he should clearly understand the goal of developing e-learning programs for the endangered language. Generally, s/he should have knowledge of language documentation and language learning. The tasks of the linguist related to this project are:

- Establishing a better language corpus for developing the e-learning programs,
- Developing pedagogical learning approaches for the endangered language,
- Developing learning objects for the e-learning programs with the help of the other two groups of participants,
- Establishing a better language archive repository.

#### 3.2 The Role of the Indigenous Community Member

The indigenous community members who participate in this project should be competent speakers of the endangered language. Basic computer skills are necessary and training in documentary linguistics is desirable so they can assist in the digital archiving. They serve as a bridge between the academic community and the indigenous community. The tasks for the indigenous members in this framework are:

- collecting and transcribing the collected multimedia language data,
- creating learning activities for the e-learning program,
- promoting the e-learning programs in the indigenous community and recruiting students for the e-learning programs.

---

<sup>3</sup> The authors led a training workshop on developing e-learning materials and on-line dictionary making for a selected group of 100 indigenous teachers, chosen by the Council of Indigenous Peoples, at National Taiwan Normal University in November 2006. The participants expressed overwhelming interests in developing their own e-learning materials following the Yami model presented there.

### 3.3 The Role of the IT Professional

The IT professionals should have knowledge of and access to state-of-the-art technologies for digital archiving and e-learning development. To serve their supportive role effectively, it is better if they also have some training in linguistics and humanities. The IT professionals are in charge of the following tasks:

- Managing the information infrastructure and the servers for the project,
- Creating the web pages and the interfaces of the e-learning programs,
- Giving support to the other groups of participants in developing digital archiving and e-learning programs,
- Creating and managing the database of the multimedia documentation,
- Maintaining a backup digital archive.

As discussed in Csató and Nathan's work (2004), the project leader should establish criteria for selection of IT professionals who are suitable for the endangered language documentation. If no suitable IT professional is available in the local community, the project leader can specify the requirements following the proposed framework and ask for help from the online linguistic community. Many computational linguists are capable of developing e-learning programs.

## 4 Conclusion

In this paper, we have proposed an integrated framework for developing an e-learning program from the research results of language documentation, using our project as an example to illustrate how we are developing an e-learning program based on the Yami language documentation project. The proposed conceptual framework includes two modules: the endangered language documentation module and the endangered language e-learning module. This same framework can be used to generate possible steps and processes for projects in other languages.

We chose not to report on technical details of the archiving work, as this has been discussed in Rau, Yang & Dong (2006), but rather focus on a higher conceptual framework. We believe a detailed explanation of the procedures leading from language documentation to the development of an e-learning program will benefit future language documentation projects. Although the current project is still ongoing, we hope the experience we have gained so far can serve as a model for other endangered language preservation projects, not only in Asia but in other parts of the world that share similar contextual resources or constraints.

## References

Bird, S. Simons, G. Huang, C.-R. (2001). The Open Language Archives Community and Asian Language Resources, *NLPRS 2001*, 31-38.

Bird, S. & Simons, G. (2003a). Extending Dublin Core Metadata to support description and discovery of language resources, *Computers and Humanities*, 37: 378-388.

Bird, S. & Simons, G. (2003b). Seven dimensions of portability for language documentation and description, *Language* 93: 557-582.

Csató, E. A. & Nathan, D. (2003). Multimedia and documentation of endangered languages, In P. Austin (Ed.) , *Language Documentation and Description*, Vol. 1, 73-84. London: SOAS.

Eisenlohr, P. (2004). Language revitalization and new technologies: Cultures of electronic mediation and the refiguring of communities, *Annual Review of Anthropology* 33: 21-45.

Fujii, S., Iwata, J., hattori, M., Iijima, M. & Mizuno, T. (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, N. P. (1998). Documentary and descriptive linguistics. *Linguistics* 36: 161-195.

Johnson, H. (2004). Language documentation and archiving, or how to build a better corpus, In P. Austin (Ed.) , *Language Documentation and Description*, Vol. 2, 140-154. London: SOAS.

Lublinskaya, M. & Sherstnova, T. (2002). Audio collections of endangered arctic languages in the Russian Federation, TSD 2002, LNAI 2448, 347-353.

Nathan, D. (2004). Developing multimedia documentation. In P. Austin (Ed.), *Language Documentation and Description*, Vol. 2, 154-168. London: SOAS.

Psutka, J., et al. (2002). Automatic transcription of Czech language oral history in the MALACH project: Resources and initial experiments, TSD 2002, LNAI 2448, 253-260.

Rau, D. V. & Yang, M.-C. (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. V., Yang, M.-C. & Dong M.-N. (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. V., Dong, M.-N., Lin, M-Y, Chang, H.-H., & Hsu, Y-C, (2005). Multimedia Materials of Yami Language, Technical Report, Department of English Language, Literature and Linguistics, Providence University.

Rau, D. V. & Dong, M.-N. (2006). *Yami texts with reference grammar and dictionary*, Language and Linguistics. Special Monograph A-10. Taipei: Institute of Linguistics, Academia Sinica.

Rau, D. V. (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).

Ward, M. & van Genabith J. (2003). CALL for endangered languages: Challenges and rewards. *Computer Assisted Language Learning* 16.2-3: 233-258.

Whaley, L. (2003). The future of native languages, *FUTURES* 35: 961-973.

Xu, S. X. (2001). *Study on language endangerment* [in Chinese]. Beijing: Central Ethnic University.

Yang, M. C. & Rau, D. V. (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, E., Yu, C.-H. & Weng, C.-X. (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, E. & Yu C.-H. (2005). Language analysis and Language Processing, Computational Linguistics and Chinese Language Processing, 10.2: 167-200.