

# An online collaborative framework for orthography system development

Sook-Kuan Chin<sup>1</sup>, Alvin W. Yeo<sup>2</sup>, Nadianatra Musa<sup>3</sup>

<sup>1</sup>Faculty of Computer Science and Information Technology,  
Universiti Malaysia Sarawak,  
Sarawak, Malaysia.

<sup>2</sup>Institute of Social Informatics and Technological Innovations,  
Universiti Malaysia Sarawak,  
Sarawak, Malaysia

<sup>3</sup>Faculty of Computer Science and Information Technology,  
Universiti Malaysia Sarawak,  
Sarawak, Malaysia.

## Abstract

This paper describes a collaborative project for orthography system development involving the Daro community and linguists. We proposed a new model which integrated with Information and Communication Technologies (ICTs) to simplify the orthography system development. Besides, collaborative work also introduced for expanding the relationship between the linguists and community in orthography system development. Currently, our project focuses on Daro community, but we anticipate that this project will extend to other indigenous communities in Sarawak as well.

This study investigates how community able to contribute their efforts in orthography system development. Hence, Daro community required to use the online platform to upload data. Their response towards the online platform was observed.

**Keywords:** *Orthography system development, collaborative work*

## 1. Introduction

Many endangered languages in Sarawak, Malaysia are not written yet. Thus, the endangered languages need to be written down to provide a standard spelling guide for the community. Orthography system is the initial step to write down the endangered languages. An orthography system is a writing system with respect to a set of symbols and writing rules [1]. However, orthography development largely difficult as it requires many costs and time from linguists. Thus, a new method was proposed to solve the current situation. The new method is applied together with Information and Communication Technologies (ICTs) to provide an online workspace for linguists and community

to conduct the orthography system development collaboratively. One of the most developed cases of the use of the Internet to preserve and strengthen an indigenous language has occurred in Hawai'i [2]. Since ICTs is an important tool to preserve the indigenous languages, thus we applied it in orthography system development to improve on the manual method.

In this study, Melanau Matu-Daro language was chosen as the development of orthography system for this language does not exist yet. Melanau Matu-Daro is a language spoken in Matu River from North Channel of Rejang River to the sea, Daro and Matu areas [3]. This study is located at Daro, Sarawak.

Melanau Matu-Daro is mostly used within the community, mainly at home or some community activities. They are able to converse with their family members or relatives but not familiar in writing.

Furthermore, so far this language does not have standard orthography. Words are pronounced and spelled based on individual's background and preferences. Based on the criteria of this language, this language had been identified as a suitable case study language to be used in this project.

Therefore, the purpose of this study is to determine that the community can contribute towards orthography system development.

This paper is organised as follows. Section 2 introduces the project framework. In Section 3 we explained the process to implement the online platform. After this, Section 4

discusses the experimental results and analysis. Finally the paper concludes in Section 5.

## 2. The Project Framework

Currently, linguists need to travel to the rural places to conduct the data collection and conduct the orthography development. Hence, the manual method does not cost effective because much time and costs were consumed to complete the orthography development. The linguists may spend a life time to produce the linguistic materials (dictionary, publication and etc) and thus some technology is necessary to help produce some useful materials [4].

To the authors' knowledge, so far there is no existing model which allows linguists to work collaboratively with the community in online platform.

Hence, we are working toward a model which will support collaborative work aimed at simplifying the orthography development. The framework for this project is applied together with Computer Supported Cooperative Work (CSCW). CSCW is a groupware that brings the people work together with the helping of computer technology (Rodden, 1991). In this context, it defines that we use computer technology to develop orthography system even the participants has different background and they are located at different places.

The framework proposes that linguists work collaboratively with the community in orthography system development using ICTs. ICTs can preserve the knowledge for the coming generation and develop a platform to access the indigenous knowledge [5]. As language represents the culture of the community, therefore it provides the important knowledge for the community. It is important to preserve the knowledge in order to pass to the next generation.

The goal of developing the framework aimed at bringing the speakers to share their knowledge about their languages. Besides, this project provided the environment for expanding a relationship between community and linguists. Linguists and community work together, each contributing their expertise to collaborate in orthography system development.

We believed that ICTs can help to increase the performance of orthography system development. Instead of travelling to the rural places, the linguists can conduct the work using online method with the community. Online communication will facilitate the exchange of information

among each others. This provides the opportunity for the local community to share the information with others.

## 3. Implementation of CODES

CODES had been implemented to provide the online working space for linguists and community to develop orthography system. The online platform implemented with PHP in Windows environment using Macromedia Dreamweaver. PHP is a widely used general purpose scripting language that is especially suited for Web development and can be embedded into HTML [7].

The online platform (CODES) consists of few major components, which include of the data collection module, IPA and spelling identification module and community evaluation module.

Data collection module allows community to upload their data once the data collected from other community members. Next, linguists use the IPA and spelling identification module to propose suitable spelling for the community. Lastly, community evaluation module allows community to express their opinion towards linguists' proposed spelling.

After the online platform being built, then it tested with different stakeholders for validating the model. The stakeholders who involved in this project are linguists and community. They used different modules in CODES to perform their action respectively.

## 4. Method

In this section, the experiment had been carried out to determine that community can contribute towards orthography system development. To realise an online working environment for linguists and community, uploading data module in CODES was used in the study. The purpose of the study was to investigate the possibility of community to contribute in orthography system development by collecting the data for linguists.

### A. Participants

The participants were two females and two males who spanned the ages of 20-80. All participants were briefed the study's purpose, method and procedure. Table 1 show the profile and information of participants who involved in the study. F1-F2 represents the younger generation who familiar with computer while P1-P2 represents the old folks who provides the pronunciation of the words.

Table 1: Participants' Information

ID	Gender	Age	Occupation	Highest Educational Background	Purpose of experiment
F1	Female	27	-	Secondary school	Collect and upload data
F2	Female	21	-	Secondary school	Collect and upload data
P1	Male	75	Retired Teacher	Primary school	Collect and upload data
P2	Female	64	Housewife	Primary school	Collect and upload data

**B. Experimental Setup and Equipment**

In this study, the 10-item Swadesh list is used. The ten words were written in Bahasa Malaysia (BM). The words written in BM is because the participants are more familiar with the Malay words. Swadesh list is a classic compilation of basic concepts for the purposes of historical-comparative linguistics [8].

Table 1 below shows the 10-item Swadesh list which used in this experiment.

Table.2: Word List

Word List	
1	I
2	wind
3	fire
4	true
5	flower
6	bite
7	fish
8	road
9	small
10	because

The equipments in this study are includes with sound recorder, timer and computer. Additionally, Internet also required in the study. Sound recorder was used to record the voices of the participants while timer used to calculate the time usage during the experimentation process. On the other hand, computer and Internet connection are necessary for the participants to access the online platform.

**C. Experimental Procedure**

The experiment involved tests with community to collect and upload data to online platform. Before beginning the experiment, we informed the participants about the

purpose of this study and experiment process.

The subjects were asked to translate the Malay words into Melanau Matu-Daro and asked the older to pronounce the word accordingly. This is mainly due to old folks of the village provides more accurate pronunciation of words. The accurate pronunciation of the words helps the linguists to identify the correct IPA and spelling. Each of the pronunciation of the words was recorded with the sound recorder. The recorded audio files were saved in computer. After this, the uploading module in CODES was used to upload the audio files. The uploaded audio files will be processed by the linguists later on. Participants had to fill out a questionnaire to observe their capability to use the online platform.

**D. Results**

Table 2 summarised that the time taken to record and upload an item less than 5 minutes. The results showed that participants help linguists to save some time in travelling because they assist in data collection process.

Table 2: Time used by the facilitator

Task	Total time used(minutes)
To record an item	< 5 minutes
To upload an item	< 5 minutes

**E. Questionnaires**

Table 3 shows the results of questionnaires. The results obtained reflected that participants felt satisfied with the online platform for data uploading process. The rating scales are from 1-5 which shows the agreement of the statements. 1 indicates that subjects very agree with the statement while 5 indicate that subject did not agree with the statement.

Table 3: Questionnaire results

1. Overall, I feel easy to use the online platform	2	1	1
2. Information in online platform is easy to understand.	2	1	1
3. I feel easy upload file to online platform	2	1	1
4. After explanation given, I can easily upload the file.	3	1	1
5. I do not face any problem whenever upload file to online platform.	3	1	1
6. I believe this online platform able to preserve the indigenous languages.	1	1	1
7. Do you will use this online platform in future time?	Ya	Ya	Ya
8. Will you recommend this online platform to others?	Ya	Ya	Ya

### F. Discussion

In general, data collection process had been carried out smoothly with the active participation of community members. The purpose of this study was to find out how community members collect and upload data by themselves. Community used the provided word list to collect and upload words using CODES. Thus, linguists able to access the data and do not need to travel to the rural places.

Besides, the results showed that the time taken is very short for them to upload data. Therefore, linguists save the time and energy to collect data while they depend on the helping from the community. In addition, they also can concentrate to create IPA and spelling for the community.

In summary, the results collected from questionnaire showed that community satisfied with the functionality of CODES. Overall, the online platform help the community to keep their language resources alive but linguists also get the benefits from it because they able to access the language resources without travelling to the field.

## 5. Conclusion

In this paper, online collaborative framework for simplifying the orthography system development is presented. The proposed model is intended to improve the performance of orthography system development. The experiment was carried out to determine that community collect data themselves help to reduce the burdens of linguists. Hence, both the linguists and community can gain a better experience in orthography system development.

### Acknowledgments

The authors would like to express the gratitude to Exploratory Research Grant Scheme from Ministry of Higher Education. And also thanks to Universiti Malaysia Sarawak (UNIMAS) for providing computer facilities as well.

Besides, the participation of Daro community members during the pilot study is greatly appreciated.

### References

1. Seifart, F., *Orthography development*. Essentials of language documentation, 2006: p. 275-299.
2. Warschauer, M. and I. De Florio-Hansen, *Multilingualism, identity, and the Internet*. Multiple identity and multilingualism, 2003: p. 155-179.
3. Lewis, M.P., *Ethnologue: Languages of the world*. Vol. 9. 2009: SIL international Dallas, TX.

4. Czaykowska-Higgins, E., *Research Models, Community Engagement, and Linguistic Fieldwork: Reflections on Working within Canadian Indigenous Communities*. Language documentation and conversation Vol 3, No1 2009.
5. Adam, L., T. James, and A. Wanjira, *Frequently asked questions about multi-stakeholder partnerships in ICTs for development*. A guide for national ICT policy animators. Published by the Association for Progressive Communications (APC), 2007.

**Sook-Kuan Chin** received B.E. degree in Software Engineering at Universiti Malaysia Sarawak (UNIMAS), Malaysia in 2009. Now she is pursuing her Master degree of Computer Science at UNIMAS. Her research interests cover the indigenous communities and language preservation method.

**Alvin W. Yeo** received the PHD degree in Computer Science at University of Waikato, New Zealand in 2002. He is now a director of Institute of Social Informatics and Technological Innovations (ISITI), UNIMAS. He has expertise in the area of Information and Communications Technology for Rural Development (ICT4RD).

**Nadianatra Musa** received the PHD degree in Computer Science at University of Tasmania, Australia in 2012. She is now a lecturer at Faculty of Computer Science and Information Technology, UNIMAS. She has expertise in IS/IT security governance, IS/IT Security Internal Controls.