Integrated Culture and Language Learning with Hakka Culture-Specific Items Annotation Platform

Chen-Chi Chang

Abstract—**Multicultural society makes the minority language less used. There is a decrease in the language proficiency of the native speakers - the younger the speaker is, the poorer his language competence. How to reverse minority language, such as Hakka language, extinction is becoming crucially important issue today. This study constructed Hakka culture-specific items annotation platform which can help to add links of culture-specific items from Hakka corpus. This system looks up the pre-build Hakka culture-specific lexicons to find the appropriate Hakka language words. This system is able to help users to produce text in Hakka written language. At the same time, this system will show the Hakka culture-specific items of the source text according to the cultural meaning. Semantic Annotation Platform for Hakka Corpus provides a solution for linking culture-specific items to concepts through various sources. This study makes these cultural resources appear as one seamless connection to related researchers.**

Index Terms—**Hakka corpus, Hakka language, minority language, semantic annotation, culture-specific items.**

I. INTRODUCTION

In recent years, there has been a dramatic proliferation of research concerned with ethnic minority language and culture preservation. For language and cultural heritage preservation research interests are getting considerable attentions. Language Revival helps people re-acquire their declining heritage languages. As researches for how endangered language revival surge past research has seen many success stories, through language learning program, even endangered languages are also available revival. Kaurna is the indigenous language of the Adelaide Plains in South Australia. Previous study presents the Indigenous language, Kaurna revitalization in settled Australia, considers the nature of the language ecology in Indigenous Australia, and advances some of the reasons for the success of language revitalization [1], [2]. Hakka language revitalization is the targeted goal of language revival efforts. It involves a reversal of language shift where people start using a language that has been moribund or threatened by extinction [3]. Hakka is a dialect of China comes from various areas with migration culture. Hakka language is the body of culture and the component of ethnicity. Multicultural society makes the minority language less used. Hakka language was not very useful for survival in society, because just a few people spoke the minority language [4]. There is a decrease in the language proficiency

of the native speakers - the younger the speaker is, the poorer his language competence. Language translation is closely related to culture. Translatability of the so called culture-specific, culturally-bound expressions has always been in the focus among theorists of translators [5]. Culture-specific items are context dependent as any kind of culture. The intercultural communication is affected by the way culture-specific item is translated [6]. The purpose of this study is to provide a solution for linking culture-specific items to concepts through various sources and help people to learn Hakka culture.

II. THEORETICAL BACKGROUND

A. Hakka Language

In search of economic opportunities, the emergence of ethnic consciousness among Hakka people in late imperial China to make their migrations. As the part of Chinese, the features of Hakka people are the historical dynamics of Hakka migrations and ethnic mobilization [7]. The amount of Hakka people is about 80 million around the world. The Hakka people have had a significant influence on the development of modern Chinese and overseas Chinese history. Hakka is one of the Chinese ethnic group and proud of its heritage and culture. Hakka people were connected by their culture, language and heritage. The Hakka people are one branch of the Chinese who lives predominantly in the provinces of Taiwan, Guangdong, Jiangxi, and Fujian in China. Many libraries, universities and Hakka cultural institutions are digitizing these documents in order to help preserve and provide open access to them. How to reverse Hakka language extinction is becoming crucially important issue today. Instead of using Hakka language words, most Hakka publications are written in dominant language - Chinese Mandarin words which can't exactly express cultural meaning of Hakka language. Today, writing in Hakka language is a difficult task because of lack of assistances by efficiency system support.

B. Hakka Corpus

Hakka corpus refers to all data collected for Hakka culture, while data set refers to all the data from the corpus that are being used for a particular analysis [8]. After decades of being overlooked, corpus evidence is becoming an important component of the teaching and learning of languages. However, John Sinclair (2004) believed corpus annotation to be a perilous activity, which somehow adversely affected the integrity of the text [9]. The fundamental problem in corpus annotation is the definition of what constitutes an entity to be tagged [10]. Therefore, the Hakka corpus in this study is based on culture-specific items.

C. Culture-Specific Items

Language translation is closely related to culture. Translatability of the so called culture-specific, culturally-bound expressions has always been in the focus among theorists of translation and translators [11]. During the internationalization, most language and culture-specific items are isolated and extracted. With culture-specific items this often seems to be impossible as the meanings that lie beyond these expressions are always strongly linked to a specific cultural context. Culture-specific items which are designed to measure concerns in one culture that may not be appropriate for all cultures under study [12]. It is not always clear which words and expressions should be considered culture specific items, even in the literature of translation several names exist for these items: realia, culturally bound items, culturally-bound, culture-specific expressions, culture bound terms/words and culture specific terms/words that signify concepts that are related to a specific culture [11].

III. RESEARCH METHODOLOGY

A. Hakka Culture-Specific Items Annotation Platform

This study constructed Hakka culture-specific items annotation platform that can help to add links of culture-specific items from Hakka corpus. The culture-specific items based on *A Chinese-English Dictionary: Hakka-Dialect as Spoken in Kwangtung Province*. The dictionary is ordered by character entries under a church Romanization that was common at the end of the nineteenth and beginning of the twentieth century [13]. This system looks up the pre-build Hakka culture-specific lexicons to find the appropriate Hakka language words. This system is able to help users to produce text in Hakka written language. At the same time, this system will show the Hakka culture-specific items of the source text according to the cultural meaning. Hakka culture-specific items annotation platform (CSIAP) supported cultural and language leaning (see Fig. 1). Automatically generated index page with all terms, sorted alphabetically, with prefix filter on top. Every culture-specific item has its own page with unique URL and Hakka culture description page. CSIAP automatically cross links all culture-specific items appearing in Hakka post and page contents
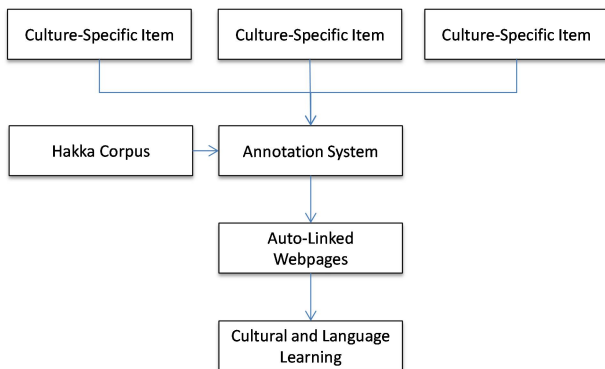

Fig. 1. System functions.

B. System Framework

This paper presents an annotation framework wherein the management process of the annotation guidelines is integrated into the annotation process. Such integration allows systematic management and reference of guidelines during annotation. For the evaluation of the proposed annotation system, we compare the conventional and proposed annotation frameworks, experiments using automatic guideline suggestion, and describe a unique feature of the integrated framework, see Fig. 2.

1) Culture-specific items management system

The Hakka language and culture experts input all the data online. Based on the open metadata standard, this system collect and process collected culture-specific items. Hakka culture-specific items included all the Hakka cultural domain knowledge.

2) Hakka corpus

Hakka Corpus in this system included Hakka knowledge base, cultural archives and culture-specific items. Hakka corpus is a collection of writings, conversations, speeches, etc., that people use to study and describe Hakka language and culture. The evaluation gives an indication of the expected quality of a corpus created by Hakka language and culture experts with relatively little training.

3) Encyclopedia models

Based on WordPress content management system, this system adapted Encyclopedia model as the core technology model.

4) Web Interface

A web interface design is a user interface that has been implemented on the Internet. This system offers online service to end users and researcher with web interface.
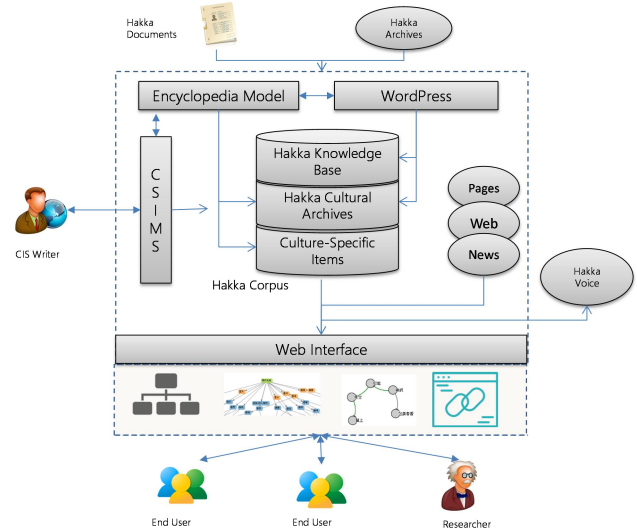

Fig. 2. System framework.

C. System Features

The features of this system are based on the idea of "cross references", the user can looks up the culture-specific items during the culture learning. Cross-link has the capability of providing culture-specific items and outsourcing process that can take language learning to the next level. Semantic Annotation Platform for Hakka Corpus integrated with culture-specific items can be the automatic link and semantic annotation solution for Hakka knowledge bases. The purpose

of this study is to provide a solution for linking culture-specific items to concepts through various sources. This system adapted the glossary WordPress premium plug-in (Encyclopedia Pro) [14] supercharged state of the art lexicon which enables the platform to create, organize and present a knowledge base with culture-specific items linking in Hakka cultural websites. Selectable content types included pages, posts, and multimedia in which the culture-specific items should be linked automatically. Hakka culture-specific items annotation platform (CSIAP) manages and organize culture-specific items as encyclopedia, lexicon, glossary, Wiki, dictionary or knowledge base. See the screen capture below outlining the main features of CSIAP and linked to the Culture-specific knowledge and resources (see Fig. 3 and Fig. 4).
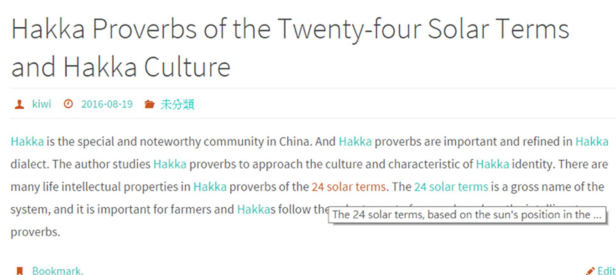

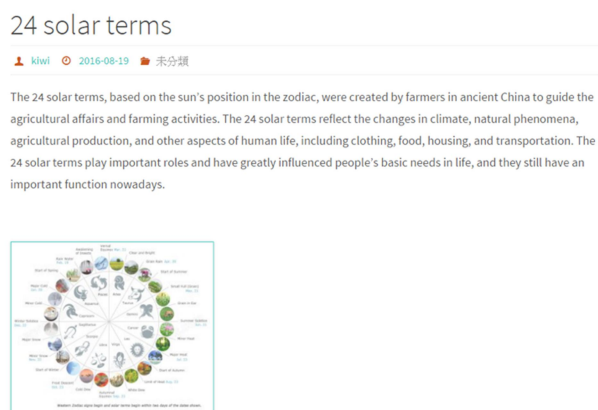Fig. 3. Culture-specific items (24 solar terms).

Fig. 4. Culture-specific knowledge and resources.

IV. DISCUSSION

A. Efficient Way to Revival the Language

This system can included additional resources, such as *Taiwan Hakka Dictionary, Hakka Words Database, the National Science Council's Taiwan e-Learning and Digital Archives Project – Digital Taiwan Hakka Villages to provide the rich Hakka knowledge base.* Hakka culture-specific items annotation platform is an attempt to make these Hakka cultural resources appear as one seamless connection to related researchers. According to statistics [15], there are approximately 7,100 languages still alive in the world, of which about 2,380 (33%) languages are in a state of loss, the loss in the case of some regions more serious, such as in North America was 92%, 47% in East Asia. Revitalization and revival of the real language can be a more systematic and efficient way. To create, organize and present a knowledge base with culture-specific items linking in Hakka cultural websites contents is the efficient way to revival the language.

B. Immersive Culture Experiences

Culture-specific items, combined with phonological information and folk literature plays an important role in the process of revival of the language. This system provided the users with immersive culture experience through the Hakka language annotation technology. The use of annotation technology that can automate a lot of the Hakka converted vocabulary of culture-specific items, save the human translation time and labor costs, nor do they need to spend a lot of resources to language experts. Annotation technology in the process that contribute to the standardization of the language, to enhance the visibility and popularity [16].

V. CONCLUSION

In this study, the integration of open source in order to implement the culture-specific items annotation platform in a short time and low cost. By constructing of full culture-specific items system, the system helps people realize more Hakka cultural knowledge. Hakka people can get more language preservation, spread, Hakka culture more fully reservations heritage. At the same time, the field of humanities scholars can build historical knowledge platform corpus. Recent years, the rapid growth of Internet resources, traditional knowledge organization tools increasingly difficult to meet the demand, many new tools have emerged, such as taxonomies, ontology, topic maps and semantic network. Culture-specific items with new web-based, the semantic structure of a particular area to be modeled, the user is able to facilitate resource discovery and retrieval of important semantic tool. Hakka culture-specific item was written so that although the Hakka are intact, but also save the Hakka knowledge. However, to through the culture-specific items to learn Hakka knowledge is not easy, because the culture-specific items compilation approach is taken in alphabetical order, if you want to be able to expand educational function, people need knowledge classification schema approach. Culture-specific items inspire readers already know from the words and then recognize that a large group of similar vocabulary. Hakka culture-specific items annotation platform, through a combination of knowledge bases and culture-specific items, with the annotation technology to help bring about a new findings in language and culture learning.

REFERENCES

[1] Walsh, M., "A case of language revitalisation in'settled' Australia," *Current Issues in Language Planning*, vol. 2, no. 2-3, pp. 251-258, 2001.

[2] R. M. Amery, *Workshop Report: Kaurna Warra Yellakaitya-Developing the Kaurna Language for Contemporary Situatio*, 2000, pp. 1-35.

[3] A. A. D. A. A. Charles and G. Isa, *The Impact of Social Mobilization Dynamics on Gbagyi Language Shift, Recovery and Maintenance,* Prospects, 2015, vol. 8, no. 2, pp. 1-10.

[4] M. H. Hsieh, *From Invisible to Visible: Stories of Taiwanese Hakka Heritage Teachers' Journeys*, Xlibris Corporation, 2012.

[5] E. Terestyényi, "Translating culture-specific items in tourism brochures," *SKASE Journal of Translation and Interpretation*, 2011, vol. 5, no. 2, pp. 13-23.

[6] A. Salehi, "Translation of Allusions in Subtitling from English into Persian," *International Journal of Comparative Literature & Translation Studies*, vol. 1, no. 1, pp. 18-31, 2013.

[7]   S.-T. Leong, T. Wright, and G. W. Skinner, *Migration and Ethnicity in Chinese History: Hakkas, Pengmin, and Their Neighbors*, Stanford University Press, 1997.

[8]   V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, 2006, vol. 3, no. 2, pp. 77-101.

[9]   J. M. Sinclair, *How to Use Corpora in Language Teaching*, vol. 12, John Benjamins Publishing, 2004.

[10]   L. Tanabe *et al*., "GENETAG: A tagged corpus for gene/protein named entity recognition," *BMC Bioinformatics*, vol. 6, no. 1, pp. 1-10, 2005.

[11]   E. Terestyényi, "Translating culture-specific Items in Tourism Brochures," *SKASE Journal of Translation and Interpretation*, vol. 5, no. 2, pp. 13-22, 2011.

[12]   R. W. Brislin, "Comparative research methodology: Cross‑cultural studies," *International Journal of Psychology*, 1976, vol. 11, no. 3, pp. 215-229.

[13]   D. MacIver and M. MacKenzie, *A Chinese-English dictionary: Hakka-dialect as spoken in Kwang-tung Province*, 1905.

[14]   D. Hoppe. (2016/5/16). *Encyclopedia Pro*. [Online]. Available: http://dennishoppe.de/en/wordpress-plugins/encyclopedia

[15]   M. Lewis *et al*., *Languages of the World,* Seventeenth edition, Dallas, Texas: SIL International, 2013.

[16]   M. Forcada, "Open source machine translation: an opportunity for minor languages," *Strategies for Developing Machine Translation for Minority Languages*, 2006, p. 1.

Chen-Chi Chang is an associate professor in the Department of Culture Creativity and Digital Marketing at National United University, Taiwan R.O.C. He received his Ph.D. degree in management and information system from National Central University in 2010. His research has appeared in *Journal of Educational Computing Research, Computers in Human Behavior, Total Quality Management and Business Excellence, Online Information Review, Journal of Scholarly Publishing*, *Aslib Proceedings, The Electronic Library,* and others. His major research interest lies in the fields of online user behavior, problem-based learning, Hakka research, electronic commerce and knowledge management.