

2010

秋風  
玉露  
洗  
清  
夢  
中  
不  
見  
人  
歸  
處  
只  
見  
白  
雲  
飛



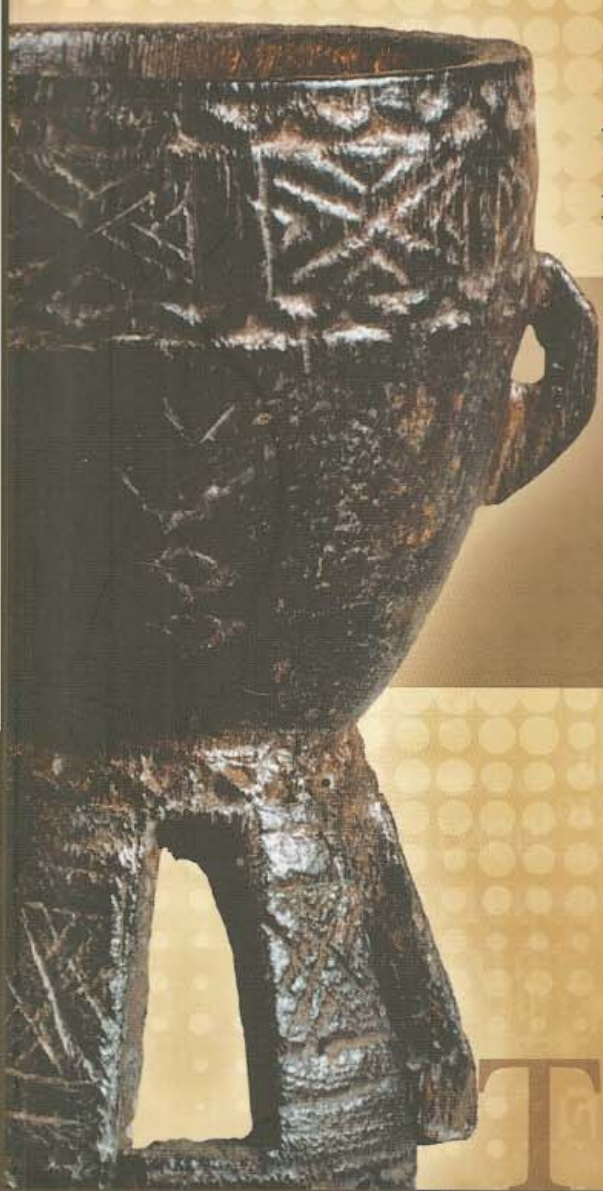
# 數位典藏與數位學習國際會議

## 2010 TELDAP International Conference

**2-4 March, 2010**

會議地點 / Venue | 中央研究院 Academia Sinica, Taipei, Taiwan

大會網址 / Website | <http://collab.teldap.tw/teldap2010/>



論文集  
Proceedings

# TELDAP



編制出版：



數位典藏與學習之海外推展暨國際合作計畫  
International Collaboration and Promotion of Taiwan e-Learning & Digital Archives Program

# Taiwan Sign Language Corpus: Digital Dictionary and Database

**James H.-Y. Tai, Jane S. Tsay**

Institute of Linguistics, National Chung Cheng University, Taiwan

## Abstract

Taiwan Sign Language (TSL) Corpus consists of two major components: a digital dictionary and a database. The TSL digital dictionary is an on-line graphic dictionary. It contains about 3000 lexical items. Each lexical item has a video demonstration of the sign with a gloss and the description of the sign in both Chinese and English. The database contains sentences, narratives, and conversations collected through various elicitation methods. These methods include: (1) native signers making up sentences or stories using key words or pictures of key words; (2) story-telling based on comic books; (3) story-telling based on the story book "Frog, where are you?" (Meyer 1969); (4) free conversations between native signers; (5) descriptions of short video clips. As linguistic theory and language universals have been constructed mainly on the basis of spoken languages, the implications of sign linguistics are far-reaching. However, sign language data are not easy to obtain. The TSL Corpus can be shared by researchers around the world. In addition to contributions in language documentation and archiving, the TSL Corpus also contributes to our understanding of universal grammar, human cognition, and origins of human languages.

**Keywords:** Taiwan sign language, corpus linguistics, on-line digital dictionary, reference grammar, sign linguistics

## I. Introduction

There are two modes by which human language can be produced and perceived. The two modes are commonly referred to as two language modalities. Spoken languages are communicated by the auditory-vocal modality while signed languages are communicated by the visual-gestural modality. Just like spoken languages around the world, signed languages are all different in different communities. Taiwan Sign Language (TSL) is the language naturally developed and used by deaf people in Taiwan. TSL is different from the sign language used in the United States (American Sign Language) or the sign language used in Japan (Japanese Sign Language or Nihon Shyuwa), even though it is historically related to Japanese Sign Language.

Lexicon and grammar constitute the two major components of a human language, regardless whether it is a spoken or signed language. With the goal of constructing lexicon and grammar of TSL, the Taiwan Sign Language Research Group have continued to build up the TSL corpus since 2001 through several research projects funded by the National Science Council in Taiwan.

TSL corpus currently has two major components: a digital dictionary and a database. TSL digital dictionary is an online graphic dictionary which contains about 3000 lexical items. Each lexical item has a video demonstration of the sign with a gloss and the description of the sign in both Chinese and English. The database contains sentences, narratives, and conversations collected through various elicitation methods. The database is necessary for the construction of a reference grammar of TSL. The online dictionary and the database together provide empirical data for linguistic and psycholinguistic research in different theoretical frameworks. In addition, the TSL corpus has pedagogical applications. For teachers in deaf education or the general public, TSL Online Dictionary can easily be accessed for learning TSL. The TSL corpus is also very useful for compiling textbooks for training teachers in deaf education. The following sections briefly describe the key aspects of TSL corpus. Since the main goal of this paper is to introduce this unique database of TSL, we do not focus on the research issues of TSL grammar. Interested readers are referred to the website of the TSL Research Group at <http://tsl.ccu.edu.tw/>.

## **II. TSL Online Dictionary**

The compilation of the TSL Online Dictionary has taken into consideration both academic and pedagogical purposes. Academically, it provides a database for those who might be interested in doing research on sign languages in general and the linguistic structure of TSL in particular, especially its phonology (basic elements for forming lexical items), morphology, and syntax. Pedagogically, for teachers in deaf education or the general public, this dictionary is a convenient and effective tool for learning TSL.

The collection of lexical items started in 2001. We have compiled a TSL wordlist based on the following sources: (i) Smith and Ting's pioneer work *Shou Neng Sheng Qiao* [Your hands can become a bridge] (Smith and Ting 1979; 1984); (ii) *Shouyu Fanyiyuan Peixun Jiaocai* [Materials for Training Sign Language Interpreters] by Department of Labor, Taipei City Government

(2002); (iii) our own numerous hours of fieldwork for the reference grammar and other research projects of the TSL research group.

The first edition of the TSL Online Dictionary was announced in 2008 which contained 1,000 lexical items, while the second edition was announced in 2009 with about 3,000 lexical items. The website of TSL Online Dictionary, also called TSL Browser, is <http://tsl.ccu.edu.tw/web/browser.htm>.

The following figure shows the homepage of TSL online dictionary.

*TSL Online Dictionary*  
台灣手語線上辭典

© 2009 Institute of Linguistics, National Chung Cheng University. All Rights Reserved.

中文	English
引用本網站資料請參照下面格式：	To cite this website:
蔡素娟、鄭浩一、李信賢、陳怡君、余瓊怡。2009。【台灣手語線上辭典】第二版。國立中正大學語言學研究所。 <a href="http://tsl.ccu.edu.tw/web/browser.htm">http://tsl.ccu.edu.tw/web/browser.htm</a>	Tsay, Jane, James H.-Y. Tai, Hsin-Hsien Lee, YiJun Chen and Qiong-Yi Yu. 2009. <i>Taiwan Sign Language Online Dictionary</i> . 2nd Edition. Institute of Linguistics, National Chung Cheng University, Taiwan. <a href="http://tsl.ccu.edu.tw/web/browser.htm">http://tsl.ccu.edu.tw/web/browser.htm</a>
蔡素娟、鄭浩一、李信賢、陳怡君、呂嘉雄。2008。【台灣手語線上辭典】。國立中正大學語言學研究所。	Tsay, Jane, James H.-Y. Tai, Hsin-Hsien Lee, YiJun Chen and Chia-Hsiung Liu. 2008. <i>Taiwan Sign Language Online Dictionary</i> . Institute of Linguistics, National Chung Cheng University, Taiwan.

62102 嘉義新巔坡鄉三興村大學路164號 中正大學語言學研究所台灣手語研究部 E-mail: [lingo@ccu.edu.tw](mailto:lingo@ccu.edu.tw) Tel: 886-5-2720411 ext. 31503

Figure 1. Homepage of TSL Online Dictionary

Each lexical item (word) has a video demonstration of the sign with a gloss and the description of the sign in both Chinese and English. The video files in this online dictionary are encoded as mov files. An example, the word SIGN LANGUAGE, is given in Figure 2.



Figure 2. Demonstration of lexical sign SIGN LANGUAGE

Some TSL signs have dialectal variations. N and S stand for the northern and southern variants, respectively. Take the word ABLE for example, ABLE\_N and ABLE\_S are the northern and southern variants, respectively. In addition to dialectal variations, there could also be free variations. For instance, the word ADDRESS has four free variants: ADDRESS\_A, ADDRESS\_B, ADDRESS\_C, and ADDRESS\_D. The variants are listed according to their frequencies, with the A variant being the most frequent, to the best of our knowledge.

### III. TSL Database

TSL database is a collection of single sentences, narratives, and conversations from a wide range of sources. The elicitation methods include: (1) native signers making up sentences or short stories using key words or pictures of key words given by the researchers; (2) story-telling based on comic books or strip cartoons; (3) story-telling using the story book "Frog, where are you?" (Meyer 1969); (4) free conversations between two native signers; (5) descriptions of short video clips.

#### 3.1 Simple Sentences

Simple sentences are collected through key words or pictures of key words provided by the researchers. The informants make up TSL sentences based on these words or pictures of the

words. The primary goal of this elicitation task is to gather information about the linguistic constituents and their relative order in a sentence, as well as other syntactic behaviors.

### **3.2 Narratives**

Narratives can be divided into two types: spontaneous narratives and elicited narratives. Similar to the collection of simple sentences, spontaneous narratives are elicited by pictures of key words given by the researchers. Informants make up short stories using these key words in TSL. Elicited narratives are collected using different eliciting materials: short video clips, comic books, or story books like the famous "Frog, where are you?" which has been used in various child language acquisition research and psycholinguistic studies for many decades. All the eliciting materials are wordless. Signers are instructed to tell the story after viewing the eliciting materials. Storytelling data are collected with the purpose of understanding the characteristics of TSL narratives.

When using video clips and strip cartoons as eliciting materials, one of the signers is instructed to tell the story of the videos and cartoons to the other signer after viewing it. The addressee is instructed to ask questions if s/he does not understand the content of the story being told. In this case, the addresser retells the story.

When using the picture book "Frog, where are you?", signers are presented with the picture book and tell the story page by page.

### **3.3 Free Conversations**

Free conversation data are collected with the purpose of analyzing linguistic structures and principles of TSL discourses. All participants in this elicitation task are acquaintances or friends prior to the elicitation and are comfortable interacting with each another. They are instructed to chat with each other freely without any given topics.

## **IV. Data Annotation**

TSL database are annotated with the aid of ELAN annotation software developed by the Max Planck Institute of Psycholinguistics in Nijmegen, The Netherlands. ELAN is specifically designed for the analysis of spoken language, sign language, and gesture, especially for the

researchers who work with multi-media corpora. The following figure shows the display of a sign in ELAN.

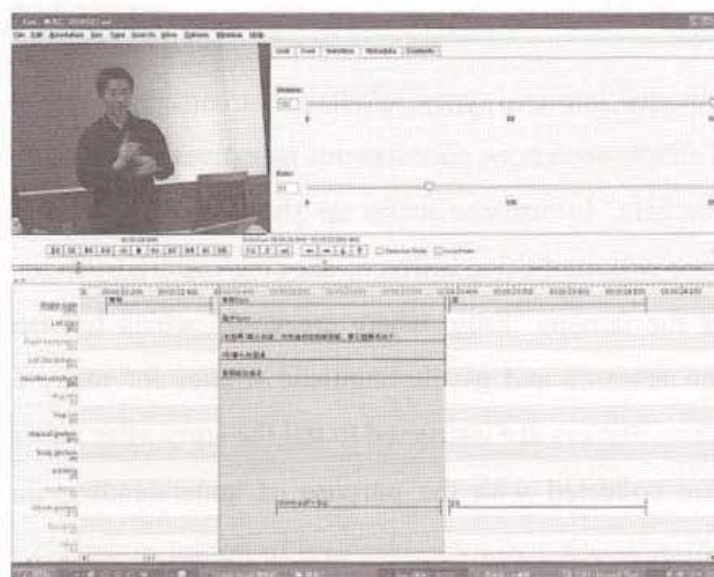


Figure 3. ELAN display of a sign

The annotations in ELAN are coded in different tiers. The annotation tiers in TSL corpus are divided into two types: manual signs and non-manual signals. The annotation tiers of manual signs include signs produced by right hand, signs produced by left hand, and classifier constructions. The annotation tiers of non-manual signals include head movements, eye gaze, mouthings, mouth gestures, manual gestures, body gestures, and body orientations.

## V. Theory, Practice, and Application

A large body of studies has demonstrated similarities or parallels in linguistic structures between spoken languages and signed languages. However, while spoken languages employ auditory-vocal modality, signed languages use visual-gestural modality. It is therefore not only desirable but also necessary to consider both signed and spoken languages in search of language universals and modality effects on lexicon and grammar in these different modalities. TSL corpus is built with the goal of studying modality effects on signed languages addressing issues including (1) documenting fundamental morphological and syntactic differences between signed and spoken languages; (2) distinguishing between modality effects and non-effects on the

linguistic structures of signed and spoken languages; (3) understanding functional differences of pointing, gesture, and facial expressions between signed languages and spoken languages.

## VI. Concluding Remarks

As linguistic theory and language universals have been constructed mainly on the basis of spoken languages, the implications of sign languages are far-reaching. The more we understand sign languages, the firmer our grasp will be of language universals. The TSL Research Group has in recent years contributed to the linguistic analysis of signed languages (e.g., Myers and Tai 2005; Tai and Tsay 2009). However, reliable sign language data are not easy to obtain. We want to continue to build up the TSL corpus to be shared with researchers around the world. In addition to contributions in language documentation and archiving for the endangered TSL, we believe that the TSL corpus would also contribute to our understanding of universal grammar, human cognition, and origins and evolution of human languages.

## References

- Department of Labor, Taipei City Government 台北市政府勞工局  
2002 Shouyu Fanyiyuan Peixun Jiaocai 手語翻譯培訓教材 [Materials for Training Sign Language Interpreters].
- Meyer, Mercer  
1969 Frog, where are you? New York: Dial Books for Young Readers.
- Smith, Wayne H., and Li-fen Ting 史文漢、丁立芬  
1979 Shou Neng Sheng Qiao 手能生橋 [Your Hans Can Become a Bridge], Vol.1.  
Taipei: Deaf Association of the Republic of China 台北: 中華民國聾人協會.  
1984 Shou Neng Sheng Qiao 手能生橋 [Your Hans Can Become a Bridge], Vol.2.  
Taipei: Deaf Association of the Republic of China 台北: 中華民國聾人協會.
- Myers, James and James H-Y. Tai (eds.)  
2005 Language and Linguistics, 6.2. Special Issue on Taiwan Sign Language.
- Tai, James H-Y. and Jane Tsay (eds.)  
2009 Taiwan Sign Language and Beyond. Chia-Yi: The Taiwan Institute for the Humanities, National Chung Cheng University.