

# Joint Optimization of Word Alignment and Epenthesis Generation for Chinese to Taiwanese Sign Synthesis

Yu-Hsien Chiu, Chung-Hsien Wu\*, Hung-Yu Su, and Chih-Jen Cheng

Department of Computer Science and Information Engineering, National Cheng Kung University  
chwu@csie.ncku.edu.tw

This article is based on the publication: Yu-Hsien Chiu, Chung-Hsien Wu, Hung-Yu Su, and Chih-Jen Cheng, "Joint Optimization of Word Alignment and Epenthesis Generation for Chinese to Taiwanese Sign Synthesis," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 29, No. 1, January 2007, pp.28~39.

Sign language is a visual/gestural language that serves as the primary means of communication for deaf individuals, just as spoken languages are used among the hearing. Deaf individuals encounter the difficulty that most hearing individuals communicate with spoken language. To improve the

communication abilities of deaf people, Alternative and Augmentative Communication technology has been adopted to develop the assistant systems for the group. Machine translation provides an ideal solution to reduce the communication barrier between these two populations. In this paper, a machine translation system is proposed to translate Chinese to Taiwanese Sign Language (TSL) and the translated TSL sequence is further synthesized by the TSL video clips to generate a Chinese text to TSL video output. Figure 1 shows the interface of the Chinese text to TSL video output system.

To develop a machine translation system, a bilingual corpus is required for training the translation models. In this paper, the teaching material of Chinese used in elementary school for the deaf children is collected for TSL sequence annotation. The annotation is completed and verified by the TSL linguists from National Chung Cheng University. A video corpus of the TSL signs is designed and filmed by considering the transitions of hand positions.

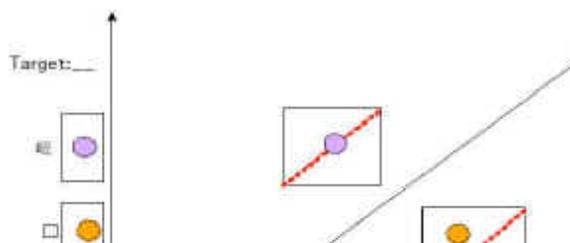
To reduce the complexity of the statistical machine translation model, a two pass translation strategy is proposed. First, the input Chinese sentence is analyzed into the primary and secondary units. Each primary fragment is composed of several secondary

Table 1 Classification of primary and secondary units

Primary Units (PUs)		Secondary Units (SUs)	
Syntactic Cluster	Examples	Syntactic Cluster	Examples
VP	喜歡(like) · 買(buy) · 有(have)	Agent	爸爸(father) · 我(I)
Negative, Positive	無法/不行(cannot) · 會/可以(can)	V (adverb)	便宜(cheap) · 漂亮(beautiful)
Conjunctions	跟/和(and) · 或者(or) · 但是/可是(but)	Dfū	很/非常(very) · 額外(extra)
Nouns (Time, proper, Nop, Mod)	明天(tomorrow) · 昨天(yesterday) · 台北(Taipei) · 餐館(restaurant) · 什麼(what) · 哪/哪裡/哪兒(where)	Nouns (common, New)	一(one) · 百(hundred) · 花(flower) · 太陽(sun) · 車子(car)



Figure 1 Interface of the Chinese text to TSL video output system



units. The first pass of translation is to align the primary fragments between Chinese and TSL. The second pass is to align the secondary units in the primary segments. Figure 2 shows an example of the proposed two-pass alignment model.

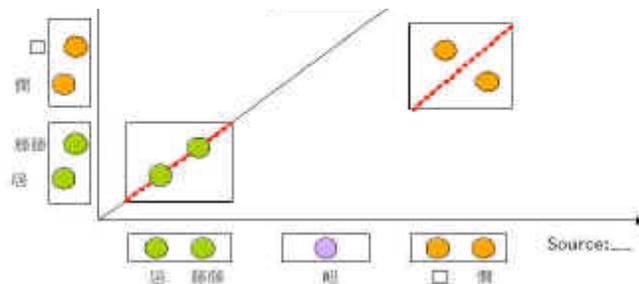


Figure 2 Search space of the alignment in translation

The translated TSL sequence is further synthesized into video output for visual communication. The video displays the translated result signed by a real human. To smooth the boundaries of video clips, the synthesis cost is defined as the difference of hand positions and hand shapes. Figure 3 illustrates the selection of suitable clips in the video corpus, and the synthesis costs are computed by the differences on hand position and moving direction of the synthesis points between the succeeding and the preceding video clips.

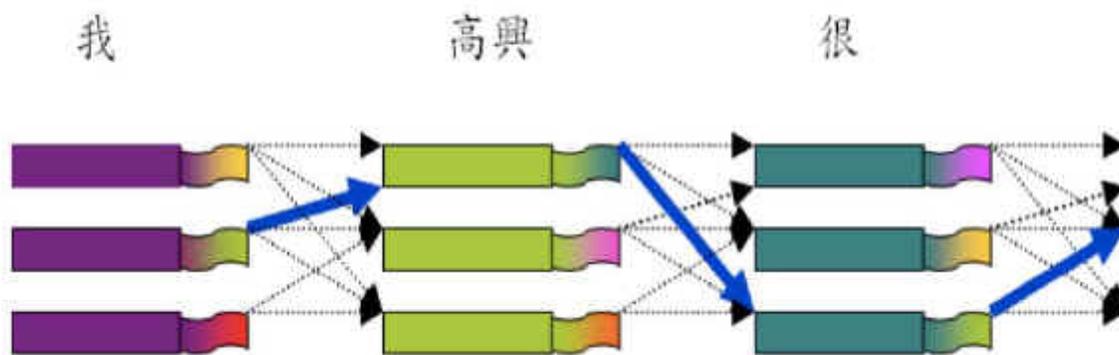


Figure 3 Selection and synthesis of video clips

For performance evaluation of the proposed MT systems, several experiments were conducted. The TOP-N translation results are adopted to evaluate the quality of translation. Figure 4 shows the correct rates for the proposed MT system and the IBM MODEL 2 system. For evaluating the video synthesis results, mean opinion score (MOS) is adopted. Ten synthesized videos with different synthesis criteria are scored by subjects using a three-level grading policy: good, fair or poor. Case A considers the position and direction, Cases B and C consider direction and position, respectively, and Case D considers no criterion. The histogram of the rated opinions is shown in Figure 5.

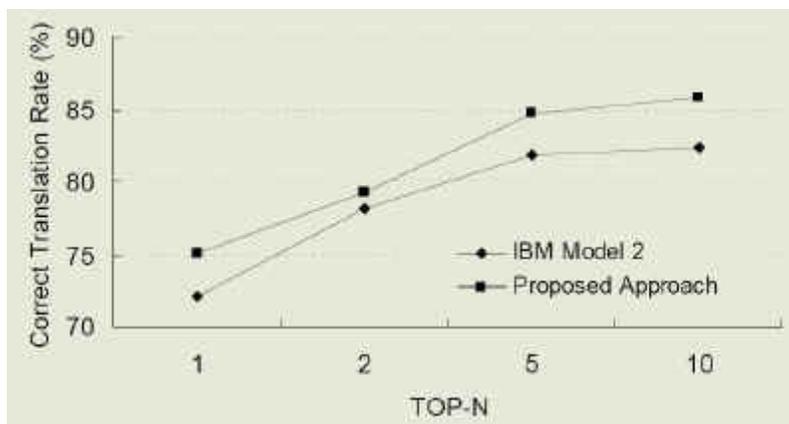


Figure 4 TOP-N translation rates for IBM model 2 and the proposed approach

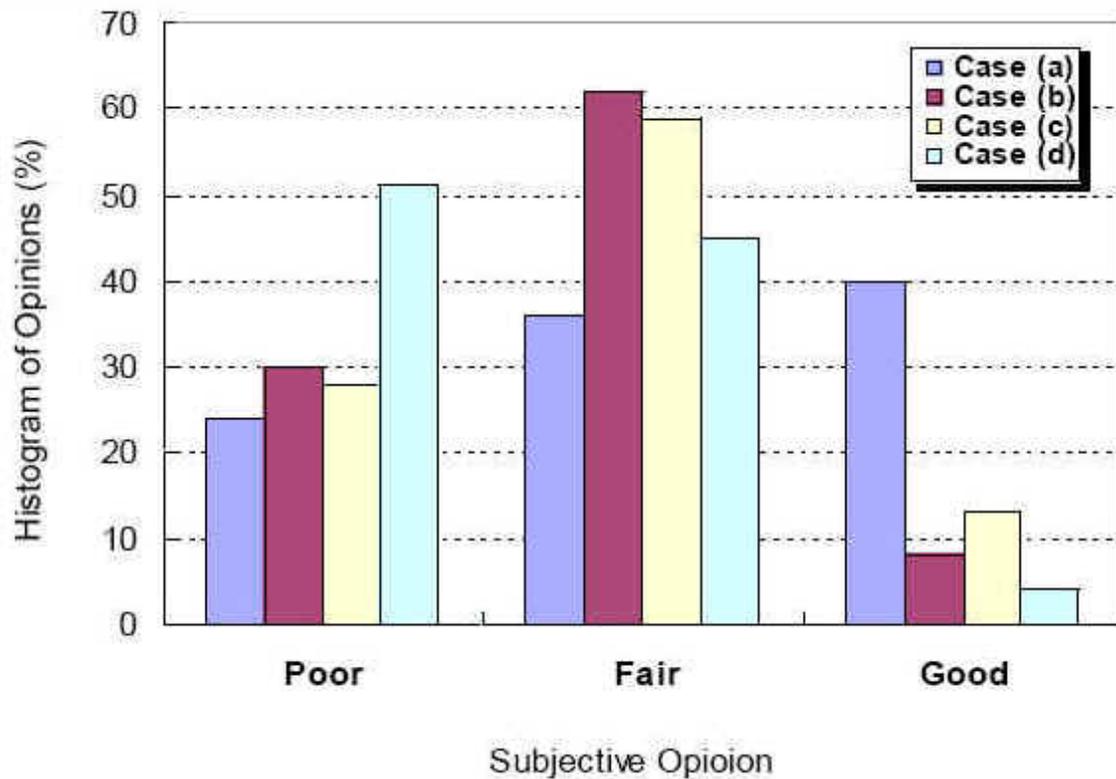


Figure 5 MOS evaluations on synthesized sign video

From the above evaluation results, the translated sign outputs with synthesized videos are promising. The technology can be integrated with educational assistants such as Computer-Aided Instruction (CAI) or TSL e-Book. In public services, the interface can be used to access the information from TV or other public machines for the deaf people.