

Transfer-Driven Machine Translation

Osamu FURUSE

ATR Interpreting Telecommunications Research Laboratories

Transfer-Driven Machine Translation (TDMT) [1, 2] is a translation technique developed as a research project at ATR Interpreting Telecommunications Research Laboratories. In TDMT, translation is performed mainly by a transfer module which applies transfer knowledge to an input sentence. Other modules, such as lexical processing, analysis, contextual processing and generation, cooperate with the transfer module to improve translation performance. This transfer-centered mechanism can achieve efficient and robust translation by making the most of the example-based framework, which calculates a semantic distance between linguistic expressions.

A TDMT prototype system is written in LISP and is demonstrated on a SUN workstation. In our TDMT demonstration, the following items are presented.

- **Target expression selection**

The most appropriate target expression is selected, according to semantic distances between input words and example words based on thesaurus hierarchy. Figure 1 shows the translation by the application of transfer knowledge at various linguistic levels and the selection of the best target expressions with the distance values.

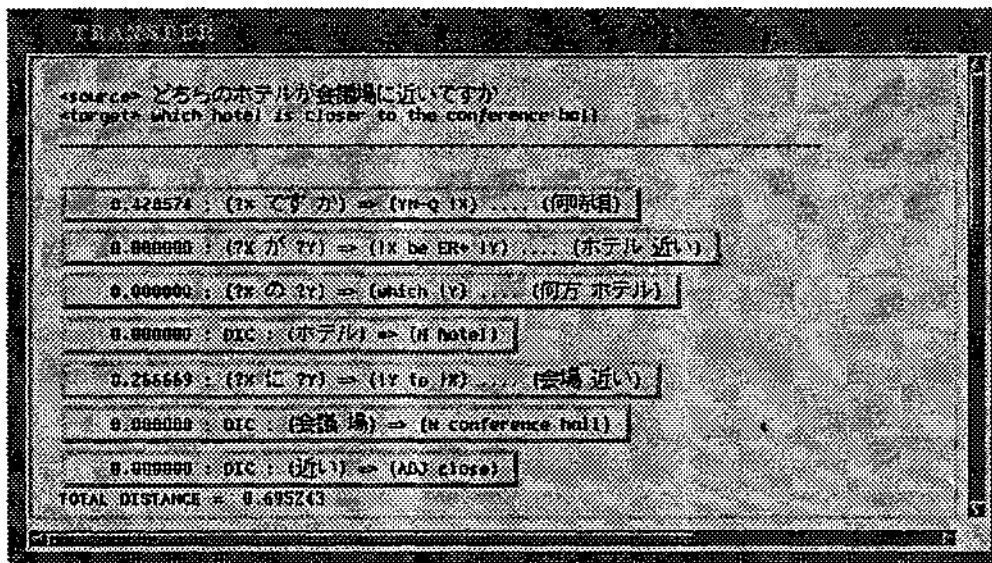


Figure 1: Target expression selection using semantic distance

- **Structural disambiguation**

Structural ambiguity sometimes brings about different translation results. The most appropriate structure is chosen by computing all possible combinations of partial translations and by selecting the combination with the smallest total semantic distance.

- **Bidirectional translation between English and Japanese**

Using bidirectional translation between English and Japanese, a dialogue between two speakers is simulated. Figure 2 shows the dialogue history, in which the left side is an English site and the right side is a Japanese site. The input sentences are displayed in black on white, while output sentences are displayed in reverse video (white on black).

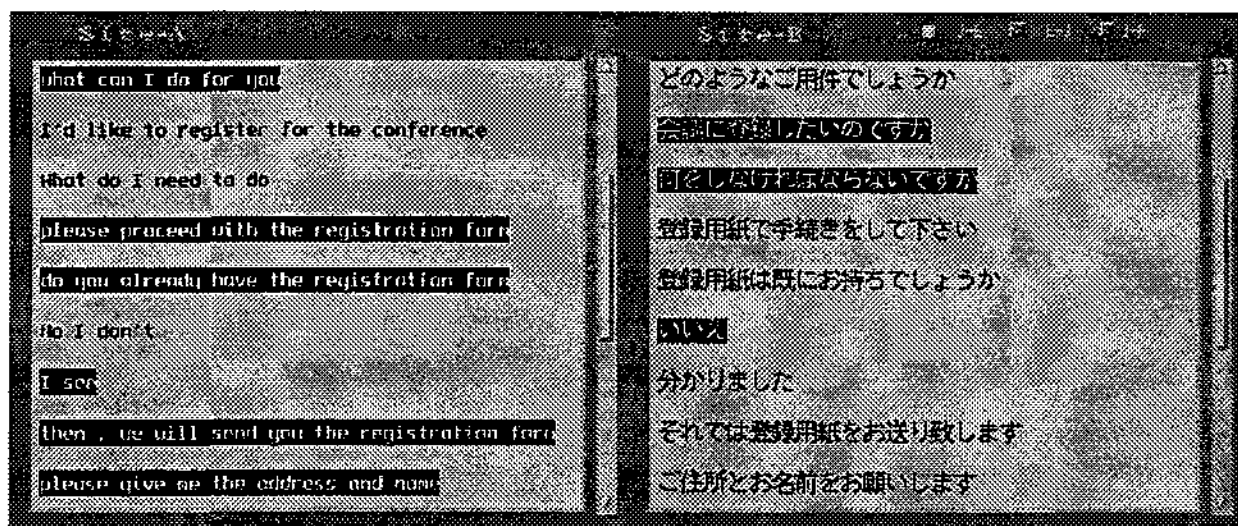


Figure 2: Bidirectional translation between English and Japanese.

So far, the system has been trained with 825 sentences for Japanese-to-English translation, 607 sentences for English-to-Japanese translation, and 225 sentences for Japanese-to-French translation, in the domain of dialogues concerning international conference registration, with a 1,500-word vocabulary.

ATR is now developing a new TDMT system which treats travel conversations with a larger vocabulary size, and plans to build a speech translation system by integrating the TDMT system and speech processing modules.

References

- [1] Furuse, O. and Iida, H. "Cooperation Between Transfer and Analysis in Example-Based Framework," *Proc. of Coling-92*, pp.645-651, 1992.
- [2] Furuse, O. and Iida, H. "Constituent Boundary Parsing for Example-Based Machine Translation," *Proc. of Coling-94*, pp.105-111, 1994.