2 • 2 Machine Translation System Using Contextual Information

Shun Ishizaki

Electrotechnical Laboratory, MITI, Japan

Japanese language is so unrelated to English that in many cases translation requires the most extensive analysis of sentences. Japanese syntax lacks much of what English syntax has. Therefore, conventional transfer approaches which use correspondence to obtain information from one parsing tree to another can hardly produce a quality translation. It is said that Japanese is rather a discourse type language than one in which parts are orderly arranged and tightly connected.

Japanese sentences often cannot be translated without referring to information contained in adjacent sentences and even paragraphs. Thus, a useful, good machine translation system must be analytical of contextual information.

In ETL, CONTRAST (CONtext TRAnSlaTor) which uses contextual information for analysis and generation are being developed.

CONTRAST uses grammatical, lexical and contextual information for analysis and generation of texts. The system must comprehend the input texts using the knowledge of background information of the subject matter stored in its concept dictionary as well as grammatical and lexical information.

Since the input text is concerned with single subject matter, every intermediate representation shows comprehension stemming from a coherent tree structure. Based on the intermediate representation, the generator determines the paragraph makeup of the text, sentence makeup of the paragraphs and styles of the sentences.

Such an advanced system does not use primitives such as those introduced by Schank. To make use of primitives would make analysis procedures complex and require an enormous amount of inferences. The granularity of concept representation adopted by CONTRAST is fine enough to allow, in most cases, to use a lexicon-to-concept conversion dictionary. Concepts usually find their corresponding lexical items in the dictionary.

Concept conversion functions is required if this method is to be used. This function is evoked when the generator is unable to find in the target language a vocabulary item matching a concept given in the intermediate expression. The concept is developed first into finer granularity (more precise concepts) and second, if necessary, to coarser granularity (more general concepts) until a correspondence between the concept and a lexical expression is obtained. This procedure seems efficient especially for multilingual translation systems such as one being developed under the ODA Machine Translation Project.

A future machine translation system may vary in function depending upon the linguistic ability of the operator. The operator with no knowledge of the target language may need a machine of almost perfect quality, whereas the operator with knowledge of both source and target languages may use his knowledge for pre- and post-work functions. Machine translation systems which use contextual information are considered most promising because they will be offering many functions to enable anyone to operate them with ease. Perhaps the ODA Machine Translation Project will provide the first chance to realize such a quality machine translation system.