Criteria for Evaluating the Linguistic Quality of Japanese-to-English MT

Satoru Ikehara
Knowledge Systems Laboratory
NTT Network Information Systems Laboratories

Knowledge of what kinds of expressions can be accurately translated is of prime importance both for potential users of a machine translation system and for people involved in MT R&D. Currently the standards suggested for the evaluation of MT quality are the ALPAC nine-level evaluation scale and, for clarity and fidelity, the Nagao two-phase evaluation standard. However, research on methods for the evaluation of translation has been limited and precedents are limited. This paper presents a method of evaluation which is based on classifying expressions that differ markedly between the source and target language and examining how well they can be translated.

The techniques needed for successful MT rely on the characteristics of the source and target languages. To achieve quality translation, it is important to have a method of translation that takes differences between the source and target languages into consideration, and evaluation factors should also be geared to the characteristics of both languages involved. For example, it is comparatively easy to determine the corresponding relationships between expressions in similar languages such as English and French or Japanese and Korean, which share common cultural backgrounds and come from the same language group. But between Japanese and English or Japanese and Chinese, which are languages from different linguistic groups, there are many differences not only in individual concepts but in the way of thinking. MT also depends upon the direction of translation.

In the case of Japanese and English, Japan imported much of her culture in earlier times from China and, since the Meiji Era, from Europe and America. Development of the forms of expression in the Japanese language has kept pace with this evolution. Therefore literal translations from these languages into Japanese can usually at least be understood. In contrast, because of the direction of the historical flow of the cultures of these countries, European and American languages lack expression structures that correspond to the Japanese language. For this reason, expressions that are peculiar to and typically Japanese are difficult to translate into English. These facts are reflected in the progress of MT development: English into Japanese MT has progressed to the stage of being put to practical use, but Japanese into English MT remains in the trial stage. In practical applications it is limited to the translation of product manuals and scientific and technical documents in which cultural differences are minimized. To get beyond this stage, it is essential to have a method for the evaluation of translation techniques that pays more attention to the differences between the Japanese and the English language.

With this concern in mind, we have developed an evaluation method that not only considers the formal differences between Japanese and English but also focuses on Japanese expressions that are peculiar to and typically Japanese. These expressions have been classified into 500 types and compiled a test set of 4,000 sentences to test the basic functions of Japanese-to-English MT.

The first step in the compilation of test factors was the classification of Japanese expressions. Based on the Constructive Process Theory of languages advanced by the Japanese linguist Motoki Tokieda, the expressions have been classified into five levels. The process whereby sentences are formed from words is divided into three stages. Objective and subjective expressions are paired up to form a set of expression elements. These units are then combined into phrase and clause structures, and the latter, in turn, are formed into sentences. Sentences are further combined into paragraphs or larger units.

The second step in the compilation considers the differences between Japanese and English. The differences were summarized in terms of six factors: fluctuations in Japanese word description, different

meanings for a single word, the nature of subjective expressions, the structure of compound words, conceptions and depictions, and the problem of articles.

The test items and the test sentences were compiled on the basis of these considerations.