〈Title〉

A MT system evaluation method

〈Authors〉

Nobuyoshi Terashima

Satoru Ikehara

Masahiro Miyazaki

Natural Language Processing Laboratory, NTT Communications and Information Processing Laboratories, Nippon Telegraph and Telephone Corporation.

〈Key Words〉

Machine Translation, Evaluation method, Japanese Newspaper, Japanese to English Translation.

〈Outline〉

Many technological problems have to be solved in order to develop a Japanese-English translation system capable of translating source text without pre-editing. To clarify these problems, an evaluation method has to be developed. However no evaluation method has yet been developed, which estimates translation quality based on translation technologies and source text characteristics.

This paper propose an evaluation method for machine translation quality. In the first place, the translation process is divided into sentence analysis phase and the conversion and generation phase. The capability of the former phase is represented by the capability of the morphemic and dependency analysis phases , and that of the latter phase is derived from the capability of English expression generation for several characteristic expressions in Japanese. An evaluation method is proposed from this point of view and is used to evaluate our MT system and also to identify the problems to develop the non-pre-editing translation system.

Specifically, 1,000 sentences of lead texts (which are typical introductory remarks summarizing a newspaper article immediately following the headlines) were taken from the Nikkei Sangyo Shinbun. Sentence structure and

characteristics in expression that would pose problems in machine translation were studied. Regarding the characteristics of sentence structuring, formal characteristics of typical newspaper articles, such as length of articles, number of "bunsetsu" (which are Japanese phrases, in most cases, consisting of a noun and "joshi" or post-positional word) and other such statistical features are studied together with the ratios of types of sentences such as simple and complex sentences. As characteristics of expression, we consider 9 types of expression may pose problems in formulation of algorithms and investigate the frequency of their appearance in newspaper articles.

Then, these characteristics of newspaper articles and quality data for each individual technology are applied to the evaluation method. Estimates of final translation quality are made and compared with actual measured values for confirming accuracy of this evaluation method and examining technical problems which remain for solution in the future.

The outline of the results is as follows.

Assuming 70% acceptability to be a practical level of quality, the number of words and bunsetsu indicate that the morphemic and dependency analysis technologies would require accuracy rates per word and per bunsetsu as high as 99.8% and 99.4%, respectively. Current technology has already reached this level. But the precision of translation for individual expressions is lower than expected values and acceptability rates for the entire sentence drops to 40% to 50%. To achieve an acceptability rate of 70%, it is necessary to achieve an average 10% improvement in all 9 types of expressions mentioned in this study and thus achieve a translation ratio of 96%.

To upgrade the translation rate of these expression, it will be necessary to establish techniques for meaning analysis of post-positional words, translation of expressions that combine noun clauses and compounded words, handling connections and embedded clauses, and verification of ellipsis and paratactical construction.