

## Japanese MT: Not a Trivial Pursuit

The level of activity in the machine translation field in Japan over the last fifteen years has been phenomenal. This was especially true during the 1980s, the heyday of AI hoopla and Japan's Fifth Generation project, when articles would regularly appear in the press announcing "90% accuracy" for the latest still experimental system. Most major electronics manufacturers had internal MT projects and a consortium was even formed to produce a telephone interpreting system. By the end of the 1980s a few commercial systems finally became available. However, since the hype far exceeded the capabilities of these systems, many professional translators, including this author, have taken great glee in ridiculing the output of the Japanese to English versions of these systems. The early systems required extensive pre- and post-editing to acquire acceptable output and the unedited output from even slightly complex sentences could be ludicrous.

To a certain extent, however, it is unfair to make fun of Japanese to English MT systems. First, Japanese text simply doesn't include enough information to translate mechanically into English. The singular/plural distinction is optional and left out in the vast majority of noun groups. Similarly, all noun groups in a Japanese sentence are optional and the subject is omitted in most sentences. Second, the usual subtle differences in meaning that plague translation between even languages as nearly identical as French and English are much worse in the case of Japanese and English. Finally, Japanese expository style is radically different from that used in contemporary English, and even a good sentence by sentence translation can sound very strange.

The results are typified by the following. "Certainly insert the stopper when you change the attachment. You will be able to get in to the machine and die when there are otherwise operation and cutting trouble of the chain." (This appeared in a factory equipment manual that required human re-translation.) Consider the first word here, which was *kanarazu* in the original Japanese. This should – of course – be rendered as "always" in this context. A pocket Japanese to English dictionary gives ten English words and several phrases as possible translations for *kanarazu*, including "inevitably," "at any cost," and "habitually." This sort of lexical difficulty is extremely common in Japanese, and often none of the words that appear in a dictionary entry will be usable. Another point here is that depending on the details of the equipment, more than one "stopper" may be required to prevent operator injury. This information (the singular/plural distinction) was not contained in the original Japanese, and the MT system arbitrarily assigned singular. Also note that the Japanese word *sutoppaa* (a loan word from English) probably should be translated as "stop" or "stop pin," depending on the shape and function of the object referred to, i.e., loan words in Japanese often have different uses than the original word in the source language and require just as much care translating as any native word.

Although J to E MT is unreasonable, some of the reasons for that unreasonableness work positively in the other direction. Since English over-specifies content relative to what is natural in Japanese, simple deletion can turn wordy output into acceptable text. Furthermore, whereas good writing in English requires that the writer make the text understandable, the tendency in Japanese writing is to put the burden on the reader. In other words, Japanese readers are willing to put up with more awkwardness and lack of clarity than English readers. (Vagueness and ambiguity are seen as virtues in academic and literary writing.) Still, the languages are different enough that even the easier direction is so problematic that Tetranet's Tetsuro Hayakawa feels that it is inconceivable that any English to Japanese MT system could ever generate even a single sentence that could be used without editing. – DJL