

Features

Improving the process of machine translation

Developers of Japanese-to-English machine translation (MT) systems face more difficulties than counterparts providing systems translating English into Japanese. "Japanese is a word-order free language," says Hitoshi Isahara, professor of computational linguistics and the president of the Asia-Pacific Association for Machine Translation. "Japanese also frequently omits the subject, so we often need some context to translate Japanese into English accurately."

And accuracy is particularly necessary for businesses selling their products overseas, which is why Isahara is working with companies in Japan's auto industry to help them provide better translated manuals for their products.

The current approach to MT most favored by researchers is to employ resource-based systems that use databases of near identical phrases in the source and target languages based on their frequency of occurrence. Isahara gives the example of the word "bank." If it is associated with "river" in a given text, then statistically it is more likely to mean the edge of a body of water, rather than an institution for lending money.

But English uses a subject-verb-object word order, while in Japanese, the verb comes at the end of the sentence. "This means we have to provide many more example sentences in Japanese, which greatly increases the size of the database, compared to when translating most European languages into English, as they also use a subject-verb-object order," says Isahara. "The computational power required for Japanese to come up with accurate matches is enormous."

Faced with such obstacles, Isahara is taking a three-step approach to improve the situation: simplifying the Japanese source text, extracting and listing salient expressions and their equivalents in a document and enhancing the post-editing process.

For example, he is devising a set of guidelines and rules for writers of Japanese manuals that will be used as the source for MT. These rules include writing shorter, simpler sentences; adding the subject when missing; and providing context when there is ambiguity.

As for post-editing, which can be costly and time-consuming, Isahara is conducting an experiment using 22 foreign students attending Toyohashi Tech to post edit machine translated versions of the university's English Web site into their own languages. The software used is Microsoft Translator.

The results of this experiment will be compared with those of post edited versions by professionals, and though Isahara doesn't expect the same degree of accuracy from the students, he notes that they have a better understanding of the context, and so this kind of collaboration could improve the post-editing process and reduce costs.

"Our approach, then, is not to focus on just one aspect of MT," says Isahara. "Rather we want to improve and support the entire machine translation process."



Hitoshi Isahara

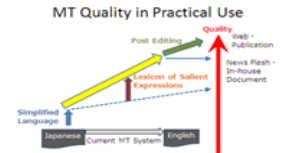


Fig.1: MT Quality in Practical Use