

## Machine Translation for the Next Century

**Jun-ichi TSUJII**

Department of Information Science  
Faculty of Science  
University of Tokyo, Tokyo JAPAN

Center for Computational Linguistics  
Department of Language Engineering  
UMIST, Manchester ENGLAND

### Abstract

The panel intends to pick up some of the issues discussed in the Summit and discuss them further in the final session from broader perspectives. Since the Summit has not even started yet, I will just enumerate in this paper a list of possible perspectives on MT that I hope are relevant to our discussion.

### 1 Nature of Translation and Limits of MT Technology

Translation is a difficult enterprise. Even a human translator with the highest qualification cannot always provide perfect translation that satisfies both readers with different purposes and backgrounds, and the authors of original texts.

Although translation seems essentially linguistic work and good command of the two languages seems sufficient, it actually requires the translator to understand source texts, which in turn requires her/him/it to have good knowledge of the subject field the source text is concerned with. More precisely, a good translator has to be able to use knowledge of the subject field to infer pieces of information lacking in the source text but nonetheless necessary for composing the target text.

Furthermore, a good human translator should be a good author of the target language as well. This may imply that she/he/it has to take into considerations what the readers would like to know by reading the texts and what background knowledge they already have.

In short, an ideal translator should be able to use diverse types of knowledge and information intelligently, not only knowledge of two languages but also that of the subject field of the text, and not only types of information explicit in the source texts but also those implicit. MT is one of the ultimate goals of Artificial Intelligence. Since we have not yet succeeded in AI in any serious sense, MT will remain far away from being perfect.

### 2 Nature of Translation and Potentials of MT Technology

Translation is a tedious work. If you have had a proper training as technical translator, translation of technical manuals becomes a tedious work after translating several manuals of a similar kind. Although one has to acquire a set of skills and should be given a set of terminology of two languages in advance, you can produce fairly good translations without actually understanding the text.

Translation becomes a repetitive work, which only requires table look-up of fixed patterns and terminology. Such a work can be carried out far better by a computer system than a human translator. A MT system can produce much more consistent translations than a human translator.

In translation of similar languages (eg. Korean and Japanese), the translation speed of human translation is limited only by the speed of typing. Although a translator has to stop from time to time and tries to understand, translation can be mostly mechanically produced. MT can surely translate texts much faster than a human translator.

In short, there are some types of texts, translation of which rarely requires human intelligence (though they may require good memory). An MT system that is useful needs not to be a product of ultimate AI technology, while it should complement the human ability.

### 3 Translation as a Cooperative Work

Translation requires cooperation among people involved. Unlike literary translation, technical translation is often carried out by a group of translators who share common styles of translation and terminology. Quality control of translation requires a systematic approach to translation that changes translation as work of individuals to that of a team. Translators with different qualifications are involved in production of single translations with different roles.

Since a translator is not a domain specialist, s/he often needs help from people who have specialist knowledge about the subject domain. In particular, direct

helps from authors of original texts are extremely useful.

In short, translation can be seen as a collaborative work that may benefit from CSCW through network.

#### **4 Translation in the Network Era**

Human translation, which MT researchers see as the ideal model of MT, has been translation of written texts, ie texts written on papers. However, the inter/intranet and WWW have changed the nature of communication drastically. Much more spontaneous and flexible modes of communication are now available, compared with communication via written texts or via phone.

Translation, either by human or by machine, has to cope with texts in the network era. Texts in electronic forms are much more versatile than texts on papers. They can carry very rich information hidden in the forms of tags like those in HGML, XML, etc., which can be used for translation purposes.

Sheer quantity of texts in WWW poses challenges to NLP/IR researchers. As rapid growth of interests in CLIR (Cross Lingual Information Retrieval) shows, MT may be integrated with other technologies such as IR, Automatic Abstraction, Information Extraction.

In short, stand-alone MT systems will not be so attractive in future, and MT modules should be easily integrated with other modules.

#### **5 Resource Sharing among MT Systems and International Cooperation**

Research and development of NLP systems in these five years have shown the importance of sharable linguistic resources. This is the case for MT as well.

However, what are sharable resources for MT is not trivial. From the view point of "MT as an ultimate goal of AI", the critical resources to be shared may be a set of language dependent concepts, as the EDR project in Japan claimed. Or it can be the Top-Level ontology which is language independent. The effectiveness (and plausibility) of these two approaches remains still to be proved.

On the other hand, from the view point of "translation as repetitive, tedious work", the critical resources to be shared may be a huge amount of corresponding linguistic expressions or translation examples. However, it seems that a mere collection of bi-lingual corpora, though useful, will not be the solution.

#### **6 Tools for Adaptation**

Our experience till now shows that there is no such a thing as a universal MT system, a system which can be used for translation of any kind. A system has to be tuned towards specific application if one would like to have reasonable quality of translation.

While to share resources may be difficult, the resources which can be adapted for specific application

would be possible. Or like the top-level ontology, resources that help to build resources for specific application would be possible as well.

Furthermore, we may need much finer modularity in a MT system in order to cope with diverse applications.

In short, what we need is a set of generic software modules and resources, the combination and adaptation of which constitutes a MT system for a given application. Competitiveness and effectiveness of such a generic system depend on, not only how good translations produced by the final application MT system are but also on how easy actual composition and adaptation of the generic system is.