Intercultural Collaboration Using Machine Translation

Toru Ishida
Kyoto University

Almost every country on Earth is engaged in some form of economic globalization, which has led to an increased need to work simultaneously in multiple cultures and a related rise in multilingual collaboration. In local communities, we can already see this trend emerging in the rising number of foreign students attending schools. Regional communities have had to solve the communication problems among teaching staffs, foreign students, and their parents, typically by focusing on relieving culture shock and its related stress with the aid of bilingual assistants. When turning our eyes to global communities, problems such as the environment, energy, population, and food require something more — mutual understanding. In both local and global cases, the ability to share information is the basis of consensus, thus language can be a barrier to intercultural collaboration.

Because there's no simple way to solve this problem, we must combine several different approaches. Teaching English to both foreign and local students is one solution in schools, but learning other languages and respecting other cultures are almost equally important. Because nobody can master all the world's languages, machine translation is a practical interim solu-

tion. Although we can't expect perfect translations, such systems can be useful when customized to suit the communities involved. To customize machine translations, however, we need to combine domain-specific and community-specific dictionaries, parallel texts with machine translators. Furthermore, to analyze input sentences to be translated, we need morphological analyzers; training machine translators with parallel texts requires dependency parsers. In the future, users might also want to use speech recognition/synthesis and gesture recognition. Even for supporting local schools, which include students from different countries, we need worldwide collaboration to generate all the necessary language services (data and software). Fortunately, Web service technologies enable us to create a workflow that assists in their creation. At Kyoto University and NICT, we've been working on the Language Grid,[1] which is an example of a service-oriented language infrastructure on the Internet.

Customized Language Environment Everywhere

Let's look at what could happen in the very near future in a typical Japanese school, where the number of Brazil-

ian, Chinese, and Korean students is rapidly increasing. Suppose the teacher says "you have cleanup duty today (あなたは今日掃除当番です)" in Japanese, meaning "it is your turn to clean the classroom today." Now imagine that some of the foreign students don't understand what she said — to figure it out, they might go to a language-barrier-free room, sit in front of a computer connected to the Internet, and watch the instructor there type the following words in Japanese on the screen: "you have cleanup duty today." The resulting translation appears as "今天是你负责打扫卫生" in Chinese, "오늘은 네가 청소 당번이야" in Korean, and "Hoje é seu plantão de limpeza" in Portuguese. "Aha!" say the kids with excited faces. One of them types in Korean, "I got it," and the translation appears in Japanese on the screen.

Is machine translation that simple to use? Several portal sites already offer some basic services, so let's challenge them with my example from the previous paragraph. Go to your favorite Web-based translation site and enter, "you have cleanup duty today" in Japanese and translate it into Korean. But let's say you're a Japanese teacher who doesn't understand Korean, so you aren't sure if the translation is correct; to test it, you might use back translation, clicking on the tabs to translate the Korean translation back into Japanese again, which yields, "you should clean the classroom today." It seems a little rude, but it might be acceptable if accompanied with a smile. Let's try translating the Chinese translation in the same way. When we back translate it into Japanese, we might get the very strange sentence, "today, you remove something to do your duty." It seems the Japanese word "cleanup duty" isn't registered in this machine translator's dictionary.

Basically, machine translators are half-products. The obvious first step is to combine a domain-specific and community-specific multilingual dictionary with machine translators. Machine-translation-mediated communication might work better in high-context multicultural communities, such as an NPO/NGO working for particular international issues. Computer scientists can help overcome language barriers by creating machine translators that generalize various language phenomena; multicultural communities can then customize and use those translators to fit their own context by composing various language services worldwide.

Issues with Machine-Translation-Mediated Communication

Even if we can create a customized language environment, we still have a problem in that most available machine translators are for English and some other language. When we need to translate Asian phrases into European languages, we must first translate them into English, then the other European language. If we use back translation to check the translation's quality, we must perform translation four times: Asian to English, English to European, and back to English and then to the original Asian language. Good translation depends on luck — for example, when we translate the Japanese word "タコ," which means octopus, into German, the back translation returns "イカ," which means squid, two totally different sushi ingredients.

The main reason for mistranslation is the lack of consistency among forward/backward translations. Different machine translators are likely to have been developed by different companies or research institutions, so they independently select words in each translation. The same problem appears in machine-translation-mediated conversation: when we reply to what a friend said, he or she might receive our words as totally different from what we actually, literally said. *Echoing*, an important tool for the ratification process in lexical entrainment (the process of agreeing on a perspective on a referent) is disrupted, and it makes it difficult to create a common ground for conversation.[2]

Even if translation quality increases, we can't solve all communication problems through translation, so we must deepen our knowledge of different cultures to reach an assured mutual understanding. For example, we can translate the Japanese term "cleanup duty" into Portuguese, but it can still puzzle students because there's no such concept in Brazil. As is well known, deep linkage of one language to another is the first step in understanding, thus we need a system that associates machine translation results with various interpretations of concepts to help us better understand different cultures. I predict that Wikipedia in particular will become a great resource for intercultural collaboration when combined with machine translators because a large portion of Wikipedia

articles will be provided in different languages and linked together.

References

1. T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," *IEEE/IPSJ Symp. Applications and the Internet* (SAINT 06), IEEE Press, 2006, pp. 96–100.
2. N. Yamashita et al., "Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation," *Proc. Int'l Conf. Human Factors in Computing Systems* (CHI 09), ACM Press, 2009, pp. 679–688.

Toru Ishida is a professor in Kyoto University's Department of Social Informatics and a leader of the NICT Language Grid project. He has been named an IEEE fellow for his contribution to autonomous agents and multiagent systems. His research interests include social information systems, digital cities, and intercultural collaboration. Ishida has a PhD in computer science from Kyoto University. Contact him at ishida@i.kyoto-u.ac.jp.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*