

Panel

MT Online: The Future is Now!

Maximizing the value of online MT. This panel addresses the principal theme of the conference, and immediately follows the opening talks by Muriel Vasconcellos, President of AMTA (on general trends in MT and the way things are going, nowadays, with the Internet), and by Mary Flanagan of CompuServe (on particular experiences CompuServe has had with MT on the Internet).

Panel members:

Elliott Macklovitch, Centre for Information Technology Innovation (moderator)
David Clements, Globalink
Mary Flanagan, CompuServe
David Greco, NetX
L. Chris Miller, Multilingual Computing Solutions
Patricia O'Neill-Brown, U.S. Department of Commerce

The Value of Online MT in the Age of the 'Cyber Society'

David Clements

Globalink

There are several uses for online MT which the public will see become increasingly available as the century nears its end. The first is online translation provided by an interactive service, such as CompuServe. In this type of MT, the service allows the user to access translation for messages to other users in special areas, or forums, and also provides a for-pay service for automatic translation of documents. So far, of the major online service providers, only CompuServe makes this functionality available. Since the competition is quite stiff between the major services (America Online and Prodigy being the other two), consumer demand and the internationalization of the online community may make MT on all the services more widely available.

Of course, no discussion of online MT can ignore the fantastic growth of the Internet and the World Wide Web. The community of users of the Web truly knows no frontiers, and the need to understand what is available in foreign-language Web sites is increasing daily. Right now, English is by far the predominate language of the Web, but as other countries develop their computer resources, a need will also arise for English speakers to rapidly translate Web pages originating in France, Germany, Brazil or Japan. There are a couple of solutions for the Internet problem. The first, as provided by Systran, is the online translation of a Web Page. Systran offers net surfers the capability of sending a URL to their server, which then translates the indicated site and sends the translation back to the

user, via e-mail. A second solution is the recent development of separate MT add-ons, such as Globalink's Web Translator, which work with popular net browser software like Netscape Navigator and Microsoft Internet Explorer. This type of add-on software supplements the value a user already has in his or her Net browser, and seems to work well for the casual users who want to use MT for information retrieval or dissemination. Right now, the Globalink products work only with four major European languages; there will be an increasing need for products spanning the range of languages, from Japanese to Russian to Swedish and Arabic. As the amount of information available on the Internet reaches 'critical mass,' MT and the Web would seem to be well-suited partners for an increasingly global future.

There will also be an increasing need for translation of other types of online documents, especially e-mail. An American who needs to do genealogy research on German ancestors may need to e-mail someone in Germany to discover resources that don't exist in the United States. MT for personal computers offers people who would have been confined to their native languages an opportunity to explore the riches of other cultures and viewpoints. So far, there hasn't been a lot of attention paid to the potential of e-mail, but many of the programs currently available would seem to be readily adaptable to this type of purpose. Many problems exist with the opportunities presented by e-mail, however. These are not unlike the problems faced by CompuServe in the translation of forum messages. People tend to be very stylistically sloppy in their casual and frequently personal writings online. Grammar, punctuation and sentence structure vary widely, depending on the individual's state of mind and the register of the communication involved. Clearly, we will need to develop more robust systems to handle the variety of styles and content present in e-mail.

Another type of online use of MT is the ability to arrange translation *jobs* through the Internet. Globalink has one such service, *Translate Direct*, wherein customers can submit documents for rapid translation, at a cost savings over having a human translator do the work. Further refinement is available, at extra cost, from human translators. This is similar to the service that CompuServe offers. Translators themselves can network with a central source to retrieve and submit jobs, which offers savings in time, as well as added choice in the availability of translations. It is also now possible for government agencies to use online MT to get rapid batch translations of large documents accomplished. With MT online, agencies won't need to maintain systems at their locations. They can use the facilities of a contractor, or other centralized agency.

The convergence of the Internet, e-mail and various online services, as well as the increasing popularity of the personal computer over the whole world, presents the greatest opportunity yet to bring MT "to the masses." No longer will machine translation be restricted to government agencies and large corporations. Nor will it be confined to individual PC users with proprietary software installed on their own hard drives. Disparate communities of interest are coming together across national borders, thanks to a "cyber society" where borders are less and less relevant. While technology has provided a bridge between countries, machine translation has the potential to facilitate the traffic on that bridge. No matter what the technology, language differences will remain. MT is the only tool which will allow the online community to assimilate, rather than be overwhelmed by, the current flood of international information and social contacts.

Online Machine Translation

Patricia O'Neill-Brown

Asia Pacific Technology Program

U.S. Department of Commerce

These are a couple of the things to consider if you are contemplating making use of online machine translation.

Character Encoding Method Issues

You must be aware of the issues involved in sending and receiving text files in languages not encoded in ASCII via the Internet. For example, there are several encoding methods for Japanese, including JIS (Japanese Industrial Standard), SJIS (Shift JIS) and EUC (Extended Unix Code). Files need to be encoded in JIS if they are to be sent and received over the Internet since it is the most stable encoding method for Internet transmission.

The Paper Problem

Many foreign language documents exist in machine readable format and are increasingly being made available for downloading via the World Wide Web these days.

However, many users have documents that they want translated only in paper form. This is particularly true in our case. At the Machine Translation Center for Japanese Science and Technology Literature at the U.S. Department of Commerce, most of the requests we receive are for Japanese patents to be translated into English. Most customers only have these documents in paper form. Indeed, it was only relatively recently (1991) that the Japanese Patent Office started requiring that all new filings be in electronic form. Patents and utility models have now been made available on CDROM in EUC encoded format; however, only those from July 1993.

If you have a need to have documents translated that are only available in paper format, it may be necessary, depending upon the language and depending upon what equipment and software you already have, to invest in software and hardware to scan the documents into machine readable code. In the case of Japanese, you can use a standard PC or Mac with enough computing power and memory, but you also have to have the Japanese (OS) operating system and Japanese Optical Character Recognition (OCR) software installed on your machine and a scanner to scan text into machine readable code.

Expertise Required by the Operator

If you are using an OCR program to scan text in, you may also need to have someone whom knows the language you are scanning in well enough to be able to make corrections to characters that have been incorrectly scanned in or to add characters that the program missed. It may be easier to correct text in alphabetic languages since there is a one-to-one correspondence between the character on the keypad and what appears on the screen (when one hits the key for "o," an "o" is what appears on the screen, for example), but for some languages, because "what you see on the keyboard is not what you want on the screen" you have to know the language to be able to input it.

Let me briefly explain the process of inputting Japanese through the keyboard to exemplify this point. Japanese characters are input via the *romaji* (romanization) to *kana* (the term used to refer to *hiragana* and *katakana* collectively) to *kanji* (Chinese character)

conversion method. In other words, to produce the *kanji* characters for the word, "Japan" on the screen, for instance, you have to know how to say it in Japanese—it is "nihon"); type "nihon" into the keyboard; it will appear either in *hiragana* or *katakana* on the screen (whichever you select); and hit a key to convert it to *kanji*.

In addition to needing a person whom knows the language to correct scanned files, you may also need to have this kind of a person if the OS environment and OCR software is in the source language and you do not know that language. For processing the Japanese language, it is possible to have a Japanese OS which has English menus (Japanese Language Kit (JLK) for the Mac), but the menus and manuals of some OSes and most commercial Japanese OCR software packages are in Japanese only.