

Online Translation: MT's New Frontier

Mary Flanagan, CompuServe, USA

Translation of online texts is a growth area for MT. This trend is driven by the increasingly multilingual environment on the Internet and the vast quantities of information that online users receive. MT can be effectively integrated with online services to produce rapid, draft translations suitable for scanning large volumes of multilingual texts.

CompuServe has deployed fully automated machine translation on several of its forum message areas, private chat rooms and in a document translation service. In each application, the goal is to provide very rapid translations that are suited for information assimilation. Translations are delivered without postediting and a high degree of acceptance of raw MT output has been achieved among users.

Because the value of online information decays quickly with time, users demand very rapid turnaround of translated documents. In addition, where casual communications such as chat, forum messages and email are concerned, the translations must be delivered at low or no cost.

Online texts, and in particular, chat pose unique challenges for machine translation. Most MT systems assume the sentence as the unit of analysis, and online texts, with their low percentage of complete sentences, present parsing difficulties for MT systems and often yield low quality output. Topic shifts, spelling and grammar errors and ellipses compound the difficulty. Automated pre-editing and dictionary development can improve output quality. From the perspective of translation production, there are also challenges in delivering real-time, high volume translations. CompuServe's chat translation prototype delivers translations in 5-8 seconds making it possible to carry on near real-time multilingual conversations. Forum messages are delivered in 2-3 minutes from the time the original is posted.

Planned MT applications at CompuServe include e-mail translation, additional forum deployments and public release of chat translation. Research areas are integration of speech recognition with chat translation and real-time translation of television closed captions.