

NHK

Do subtitling and MT mix?

One of the most intriguing applications of MT in Japan can be found at NHK, the Japanese Broadcasting Corporation. As part of an ongoing search to improve the speed and efficiency of translating foreign video feeds for Japanese news programs, NHK has extensively adapted and extended Catena's star English-Japanese system for its own purposes (the system is now simply called the NHK MT system) and, since 1989, has been using it on an experimental basis (52,000 words in 1992) to translate a small segment of English-language television news every day. In the current configuration, a translator transcribes the spoken English, shortening and simplifying where necessary, runs it through a spellchecker, and feeds it, sentence by sentence, to the NHK system. The translator then post-edits the output where necessary and the output is re-synchronized with the video. While the original NHK system ran on a Sun 3, the NHK has recently developed a highly attractive Macintosh-based package for transcription and translation. Running on a powerful Mac Centris 650, this integrated system allows onscreen viewing of video, which facilitates the arduous synchronization process. According to NHK's Hideki Tanaka, the system does not increase overall throughput substantially transcription of the English remains a difficult, laborious process and good transcribers are rare and expensive in Japan but the NHK remains committed to experimenting with the system. However, a major breakthrough may be on the horizon; US broadcasters may soon be offering "closed caption" services to foreign broadcasters like the NHK. If and when this occurs, this could dramatically increase the possibilities for automating the subtitling process and NHK's long investment in MT may well pay off.

The NHK also deploys its MT system for one of those classic domains for which like weather reports and avalanche warnings MT is so well suited. This is the translation in real-time of the Associated Press Economic News for internal use within the organization. In contrast to news broadcasts, which are virtually unlimited in their subject matter, the Economic News is a narrow domain with a fairly limited range of expressions ("Other late dollar rates in Europe, compared to late Monday...") and the NHK is able to exploit this fact successfully. To supplement the NHK's basic lexicon of 120,000 entries, 400 fixed patterns have been coded into the AP system, and these cover nearly thirty percent of the incoming material, offering an accuracy rate of one hundred percent. If the system is not able to translate a piece of text on the basis of these patterns, it falls back to its domain-specific grammar, which has an average accuracy rate of seventy percent. If this fails, it in turn falls back on the general grammar, which averages only ten percent accuracy. The output is of high enough quality for internal use without post-editing.

NHK, 1-10-11 Kinuta, Setagaya-ku, Tokyo 157, Japan; Tel +81 3 5494 2314, Fax +81 3 5494 2309