Are We There Yet?

Road-testing the latest in Japanese/English machine translation

In an occasional series of new product reviews, we asked Peter O'Rourke at Mitaka, a European/Asian-language service provider, to put Neocor's Typhoon and Tsunami Japanese/English translation systems through their paces in a real-life work environment.

We at Mitaka have been using Japanese machine translation for the last half-dozen years. When we began assessing the Typhoon Japanese-into-English MT system and the Tsunami English-into-Japanese system, we were naturally interested in the usual criteria (difficulty of installation, ease of use, accuracy, etc.), but we also had a number of other reasonable expectations based on our previous experience.

By "reasonable," we mean knowing that life with an MT system is not always peaches and cream, after that initial thrill of unpacking the box and the novelty of encountering a new system. The problems tend to surface only after the first example of perversity, or when eccentricities appear for which the manual does not prepare you. In other words, we came armed with very realistic expectations of both the potential strengths and weaknesses of machine-translation systems.

We viewed this test as a combined enterprise involving both our IT department and the production-management team. An MT system is a working tool, and we wanted to avoid testing it in a vacuum. It was vital both to address the needs of production staff with no specific IT background, and to stimulate the interest of an IT staff who saw it merely as an extension of earlier development work rather than a potential cost at the production level.

Hurry Up and Wait

The team that attempted to install the software using the manual all agreed that this was straightforward. They ranged from people with considerable IT and MT experience to complete neophytes in both fields. Start up was thus a success.

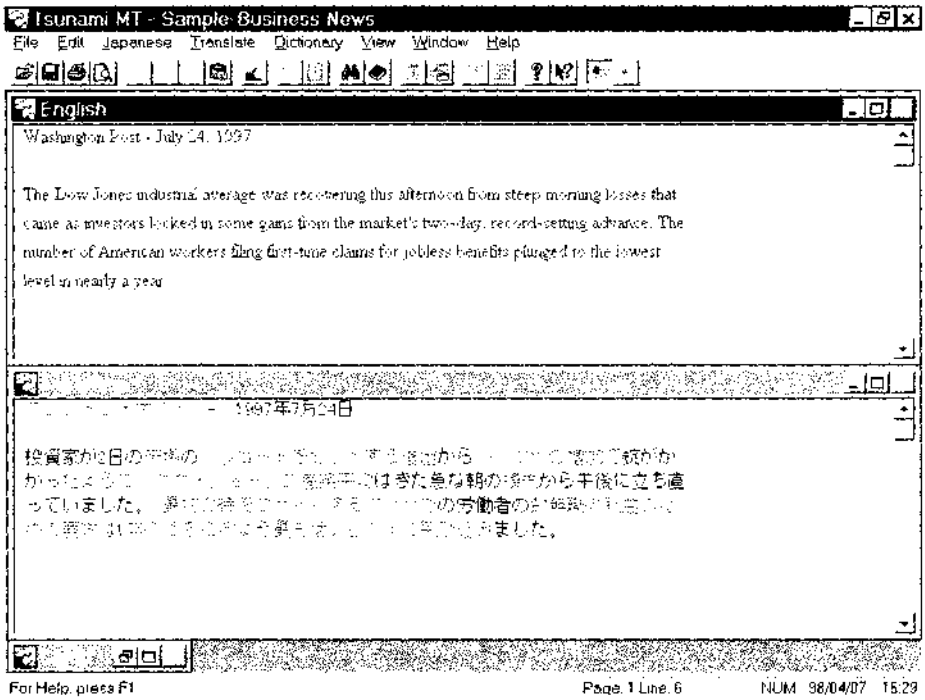
To test the OCR capability, we used standard Twain-compatible scanners. We began by using the test texts which came with the software, but then changed to text that had been generated in-house. Regardless of point size and resolution (dpi), the results were disappointingly indifferent. Of more concern, however, was the absence of consistency with the same text, giving varying levels of accuracy when the identical copy was put through on more than one occasion. In one case a highly experienced MT user encountered rates of 30-percent and 90-percent accuracy with consecutive passes of the same text!

Based on our testing, we calculated that the maximum we could consistently expect from the OCR facility was little over 30 percent. We understand that this figure is far below that expected from OCR when operated by experienced staff. If output accuracy levels fell this low, time spent proofing and checking would negate any value it might offer.

For the subsequent testing, we relied upon text that had already been saved electronically, thus skirting the program's perceived OCR bottleneck.

Neocor Typhoon Japanese-into-English MT

Although we regularly translate significant volumes of Japanese into English, we have always done this manually. We have test-



Feature	Assessment
Installation	Very straightforward
Ease of Use	High
OCR Feature	Disappointing
Components in Package	Three independent systems(OCR, E-J, J-E). There is some communication between the OCR system and the translation systems, but each system is a self-contained unit. Additional email package with multilingual capabilities. Tsunami and Typhoon also include a Japanese FEP (front-end processor) for Windows and both support OLE so that Japanese files can be inserted into standard Windows packages.
Pre/post-Edit Requirements	Works best where pre- and post-editing facilities are available. Pre-editing the text will improve the accuracy of the translation and allow subject-specific words to be added to the dictionaries before translation commences.
Translation Method	Sentence by sentence. The layout of the "before" and "after" screens can be changed.
Dictionary Encoding	Straightforward for adding individual new words to the dictionary. Easy-to-use built-in tools. Two user-defined dictionaries allowed. No method for batch-entering user-defined word lists.

driven most of the standard MT packages for this language combination (Globalink, Systran, Logos, etc.) but have never seriously contemplated becoming active users.

For this test, we used several short business letters and parts of a management manual we had translated for a client and for which we already had electronic versions in English. None of these tests was pre-edited. We felt that if we had to spend time pre-editing short texts, we might as well complete the translation manually.

Results were varied, but all versions allowed us to gain sufficient sense from the output for them to be considered a success, given the (admittedly debatable) assumption that the prime purpose of MT packages, not dedicated to a specific genre of translation, is to provide quick, albeit "down-and-dirty," translations.

Neocor Tsunami English-into-Japanese MT

Our initial testing methodology was the same as that used when quantifying the performance of our in-house Japanese MT system. This entailed locating two small English-into-Japanese texts, and translating and editing them normally via some of our in-house staff translators. We then ran the same jobs through the MT system using another Japanese staff member who had not been involved in the manual translation. The various test formats are: raw text; text with pre-editing only; text with post-editing only; and text with both pre- and post-editing.

We apply the same tests using our own equipment and compare the results, times, etc. This ensures a realistic norm against which to measure the results. However, given the time spent developing and customizing our own system, as well as its glossaries and dictionaries, we decided that comparative testing was stacking the deck against the newcomers.

When working with Tsunami on letters and short business texts, we got the results we expected. The overall meaning was comprehensible. Longer texts tended to deliver poorer results. Small amounts of pre-editing, post-editing and pre- and post-editing gave the improved results expected for both long and short texts.

Bear in mind that we were testing the Neocor packages in an environment familiar to us, but in which we would not normally use MT. On the basis of our experience, Tsunami performed at least as well as we expected, and there is every reason to assume that it would be an extremely interesting package, once fully fine-tuned to our regular course of business.