[*Language Industry Monitor* 17, September-October 1993]

Toshiba

With ASTRANSAC, Toshiba got it right: connectable, compatible, and configurable. Where to now?

Having launched the Japanese wordprocessor revolution in the late 1970s, it was only natural that Toshiba should join the MT race. And, indeed, 1985 saw the introduction of the Japanese-English ASTRANSAC followed by an English-Japanese version five years later. As Shin-Ya Amano, a Chief Research Scientist at Toshiba's Kawasaki R&D Center, points out, Toshiba has been involved in NLP research for many years, since 1971 to be exact. Back then, NLP efforts were hampered by the lack of powerful computers, but today that is no longer a problem. ASTRANSAC runs on Sun workstations and Toshiba's compact SPARC-based laptops. The latter may be small but they pack plenty of power: up to sixty mips. While these laptops may be portable and self-contained, the ASTRANSAC software is not designed to be an island in the office. The package includes built-in OCR software which uses the system's MT dictionaries to improve recognition performance. In the all-important file compatibility department, ASTRANSAC can process FrameMaker documents.

While the first generation of Japanese mainframe-based MT systems may have been lacking in interactive qualities, Toshiba has expended considerable effort in refining the user interface of ASTRANSAC, currently implemented in OpenLook. Like similar systems, ASTRANSAC displays source and target texts aligned in parallel, sentence by sentence, in its multilingual text editor. If operated in interactive mode, the system will highlight ambiguities in the target text, such as the notorious prepositional phrase attachments, and the user can view the different readings.

The Toshiba MT group has enhanced ASTRANSAC with custom-iz-ation facilities that allow users to "personalize" the translation output by adjusting various default values. Going from English to Japanese, you can choose "polite" or "normal" style, "you" omission, and the default trans- lation for participial constructions: sequenced chrono- logically, causally, or temporally. Going from Japanese to English, you can choose between the passive or the imperative voice for subjectless sentences in Japanese sentences (quite common), default articles and determiners, and default tenses for verbs.

ASTRANSAC has a basic lexicon of 50,000, technical lexicons for several fields, and space for up to 200,000 terms in its user dictionary. According to Amano, the ASTRANSAC basic lexicon has been fine-tuned with 100,000 lexical rules. Toshiba's ASTRANSAC is a mature, well-established MT system by any standard and it enjoys a good reputation in the MT world. There are an estimated three hundred ASTRANSAC licenses (mostly E-J) at two hundred sites in Japan, such as Mitsui and Company (a large trading company) and the Japanese subsidiary of Unisys. ASTRANSAC sales and marketing is handled by a department in a remote part of the company, and Amano and his team seldom have contact with users now.

While the development work is completed and the system now enjoys a solid base of users, this does not imply that there is no room for improvement. Customization facilities and lexical rules go part way to making the system more exible and help broaden its coverage, but ASTRANSAC, like other comparable MT systems, has reached the upper boundaries of what the current tech- nology permits. Amano suggests that they are at a stage where there is not much more they can do to the existing system and new techniques are needed if they want to improve its output quality or expand the number of domains the system can handle. Amano also points out that many of the problems they would like to solve are not grammatical. Such common phenomena as ellipsis, apposition, coordination, and insertion are difficult pragmatic issues which computational linguists have yet to solve. They are stumbling blocks on the path to better systems. Even the use of articles and determiners, which trips up Japanese MT systems, is, strictly speaking, not

a grammatical issue but one of world knowledge.

As part of their on-going research efforts, Amano and his colleagues have recently been looking into example-based techniques to try to improve the output of the system, possibly at the transfer stage. This approach requires a large bilingual corpus and they have been gathering materials from another division of Toshiba, a large-scale integrated chip producer. The latter has been supplying several hundred pages of documentation per month, the Japanese original and the (theoretically) corresponding English translation, for the MT researchers to work with. For Amano and his colleagues at Toshiba, not to mention other Japanese MT teams, this is new terrain; it is still a long way from being tried and proven. The way forward remains uncertain.

Toshiba Corp., R&D Center, Communications and Information Systems Laboratories, 1, Komukai Toshiba-cho, Saiwai Ku, Kawasaki, 210 Japan; Tel +81 44 549 22 39, Fax +81 44 549 22 63