

TONY WHITECOMB

REPORTS ON DEVELOPMENTS FULL OF FAR EASTERN PROMISE

JAPAN'S CHANCES IN THE GREAT MACHINE TRANSLATION

STAKES

Though still in his forties, Dr Kosho Shudo belongs to the earlier generation of Japanese MT researchers. His activities date back to 1968 when he became involved in pioneering an experimental Japanese-English MT system at Fukuoka University for translating material about transistor circuits. By 1974, this system used a set of 4,000 rules for semantic processing which could reportedly disambiguate 1000 ambiguous words. Now professor at Fukuoka University's Department of Electronics Engineering, Shudo spent a year in 1981 working on the METAL MT system under Jonathan Slocum in Texas. Shudo specialized in the MT treatment of idioms and collocations, and he has compiled a Japanese idiom dictionary for MT purposes.

I met Professor Shudo at the *Analogy-Based Machine Translation seminar* in Kyoto last summer. Apart from providing welcome protection against the high humidity and 36°C August temperatures, the ATR (Advanced Telecommunications Research) Institute proved the perfect forum for exchanging views on current MT developments throughout the world.

Asked to compare Japanese MT to advances in the West, Shudo predicted future Japanese supremacy, but for reasons that non-Japanese might find surprising. The Japanese computer industry is coming to grips with the man-machine communication problem and the race is on for an intelligent wordprocessor which will handle the interface using natural language. NEC, Toshiba, Hitachi, and the rest are currently competing to add some form of linguistic-semantic intelligence into their *wapuros*. Advances in this key area have been small but highly significant. Wapuros now, for example, use the popular method of kana-kanji conversion, allowing the normal keyboard to be used for input tasks. This involves typing Japanese words using a pronunciation template. Since each syllable in the language corresponds to several kanji (Chinese-like) characters, language-statistical and even more advanced syntactic, semantic, and even discourse knowledge is needed for the system to select the right typographic choice from the equivalent-sounding input formats. Last year, Toshiba announced the development of a neural network chip to handle this kind of processing. But even then, inputting Japanese text using a more-or-less standard keyboard will still remain a far more interactive process than with alphabetic languages, since the user will be responsible for overriding the computer's first choice of kanji character if it proves to be the wrong one. Shudo points out that the ongoing cycles of word-processor upgrading are generating the cash flow, while at the same time new-generation Japanese MT know-how is steadily advancing in the labs of the same computer-manufacturers.

This incremental scenario, whereby local progress in putting language awareness into wordprocessing feeds forward into MT applications, is quite different from what has been happening in the West, where there is a profound gap between work on wordprocessing (however many bells and whistles packages might feature) and a fully-fledged MT system. Certainly, there have been 'intermediate' solutions, such as Claude Bédard's MPT, Melby's Mercury, and the ALPS and WCC CAT products, and now other PC-based systems, but these address only a very narrow band of the whole wordprocessing market—and they make very little investment in semantic processing technology. Leading wordprocessing software developers like WordPerfect and Microsoft are not (yet?) waging battles over language/semantics issues. Spell and grammar checkers are considered simple utilities, whereas DTP facilities for mixing text and graphics have proved far sexier accoutrements. As has already been predicted, it looks as if Japan's Achilles' Heel—the computer-unfriendly nature of the culture's writing system—may be the very source of NLP design strength in the future.

MT researcher Tony Whitecomb has just returned from six months in Japan.