

Is Japanese Text Processing Peculiar to Japanese?

MASAAKI KUROSU *

Japanese text processing research has undergone dynamic development during the decade since the appearance of the first Japanese word processor on the market in the late 70s, and has yielded substantial results. The purpose of this special issue is to review the recent history and state of the art of research, and to preview future developments.

As may easily be imagined, many problems in this field of research are due to the characteristics of Japanese as a language. The following are typical peculiarities of Japanese: (1) there are more than 3000 Kanji characters in common use, as well as Hiragana, Katakana, and alphanumeric characters, (2) the same character may have several graphical forms, (3) the ability to input this vast number of characters requires special hardware and software, (4) the writing system is bidirectional, horizontal and vertical, and (5) spaces are not used between words.

Because most concepts and devices in information processing originated in the U.S.A., Japanese researchers have had to adapt them to the Japanese language and culture. Their efforts to solve the problems inherent in Japanese may, unexpectedly, have some universal significance. If an information processing system is to become truly universal, it must be reconciled with various cultural variations, including the Chinese language, which has far more Kanji characters than Japanese, and Arabic, which has a right-to-left writing system. I think that the history of the Japanese text processing research has been a huge experiment toward such a reconciliation.

In this issue, six articles provide a review of the experiment and the future of the technology. The first, by Akira Miyazawa, discusses the character set and character code of Japanese, of which I have already mentioned some of the problems. The subsequent four articles deal with the theme of input. Tsutomu Kawada reviews how Kana-Kanji conversion was developed as a keyboard-type input method, and discusses the human interface in Kana-Kanji conversion systems and current applications for proofreading support and automatic formatting. Hajime Ohiwa and Hiroshi Tatuoka emphasize the importance of Popularizing blind touch-typing in Japan, and propose the standardization of the interface between the device and the keyboard, and not of the keyboard itself. Masaki Nakagawa discusses the on-line recognition of handwritten characters as the most promising non-keyboard input method, and makes an extensive review of the technology. A report by Hajime Ohiwa et al. deals with the first input phase of Japanese text processing using an idea processor or the computerised KJ-method, which was originally a chart-based idea generation method. In the final article, Tetsuo Tamai overviews Japanese-based specification languages and Japanese-based programming languages as applications of text processing, with some experimental results.

To our regret, this issue lacks an article dealing with the output of Japanese text. Desk Top Publishing in Japan entails many difficulties deriving from the intrinsic nature of Japanese that arise when it is used with a foreign language such as English. We refer the reader to papers appearing in regular issues to the Journal of the Information Processing Society of Japan.

*Design Center, Hitachi Ltd.