Chinese-English Machine Translation

Project on Linguistic Analysis

University of California

Berkeley 94720

Research on machine translation from Chinese to English under the direction of William S-Y Wang was carried on at the project on Linguistic Analysis (University of California, Berkeley) during the period 1967 to 1975. During the early part of the effort, System I was developed which includes: a) CHIDIC: A Chinese to English machine dictionary of about 80,000 entries (60 percent physics, 30 percent biochemistry, and 10 percent general), and b) Monolithic grammar of about 4,000 rules (context-3, phrase-structure rules). In 1973, two factors caused redesign of the approach toward the development of System II. One, the grammar had become so cumbersome and ad hoc that its effectiveness as well as its potential for improvement were curtailed. Second, the sponsor requested conversion of the system from CDC machines to IBM machines. In response to these factors, System II is designed along the lines of "structured programming" (i.e., it is built on self-contained program modules). It is also designed to be machine-independent, so that it can be implemented at different computer installations.

Efforts in research and development have been aimed at an operational system. We have experimented with numerous trial sentences as well as several "live" texts (from articles of 3,000 characters in length) and have accumulated machine texts of over 560,000 characters. System II is incomplete, lacking especially the machine-editing of output to conform to those morphological features absent in Chinese but required in English.

William S-Y. Wang

Professor of Linguistics, University of California, Berkeley

Wang received his Ph.D in Linguistics at the University of Michigan in 1960, and was appointed Professor of Linguistics at the University of California (Berkeley) in 1967. He is interested in the structure and function of language, including the processes whereby one language is translated into another. Some of his work have been on system simulation of linguistic processes humans do easily, such as speech recognition and machine text analysis. He is the editor of a bilingual journal, Journal of Chinese Linguistics.