

EXPERIMENTS WITH ODD LANGUAGES

Eric Grinstead

Asian languages may seem exotic, with the special difficulties of strange cultural vocabulary, script, and the few opportunities of communicating in them. However, Chinese, Arabic, Indonesian, Hindi, and Persian are not only languages with long cultural histories, but also very influential *linguae francae* in their political and cultural spheres today.

The problems in Chinese seem to be the reverse of those studied by Europeans. One begins with basic roots, neatly separated, and a "machine translation" must show how they are put together. The Chinese written language, being composed of unit areas, has always been much better organised grammatically, phonetically, and bibliographically.

The line-plotter has been used for Chinese since 1968, at least, and now we have the data screen, the matrix printer, jet plotters, and, in Japan, the line-printer. There is teaching in the University of Illinois. Coding in four-digit numbers is well established, though the usual goal, to input all available texts, would be too much for foreign students to attempt, but could be easily achieved by Chinese students.

I am interested in: (1) Design of special vocabularies to be put into a small computer, in compact Chinese and English.

(2) A database for finding quotations, built on the existing concordances and indexes. Two rather rare characters could well establish a unique original source. This is an attempt to bypass the dictionary, and go to the original text, for which there is often a standard English or French translation.

(3) Index to the second character of a two-character compound. This is to speed up technical translation by reducing look-up time.

Alphabetic scripts present few problems. Golfballs are available for most languages already. Here I am interested in inputting the basic dictionary, and finding some algorithm that will enable the text to be compacted as much as possible for use in small desk computers.