[From: *Translating and the Computer*, B.M. Snell (ed.), North-Holland Publishing Company, 1979]

AN INTERACTIVE ON-LINE MACHINE TRANSLATION SYSTEM (CHINESE INTO ENGLISH)

Shiu-Chang LOH and Luan KONG
Hung On-To Research Centre for Machine Translation
The Chinese University of Hong Kong

ABSTRACT

The present on-line system is a direct conversion of the CULT (batch) machine translation system which has been used since January 1975 on a regular basis in translating Chinese mathematics journals. The pre-editing procedures used in CULT are being implemented in the present on-line system by means of editing programs. The enormous problem of inputting Chinese texts is being solved by keying the text directly with a newly designed Chinese keyboard.

Introduction

This paper describes briefly an interactive on-line machine translation system capable of translating Chinese into readable English, which has recently been implemented at the Chinese University of Hong Kong.

The design of this on-line system is mainly based on the algorithm developed for the much used CULT (Chinese University Language Translator) system which was first developed with the IBM 1130 system in 1969 and then implemented on the ICL 1904A system in 1972 (Loh 1972). CULT has been used, on a regular basis, to translate the Chinese mathematics journal, *Acta Mathematica Sinica,* since 1975 (Loh and Kong 1977). The capability of CULT has also been further developed to assist translation in both directions, that is CULT is now capable of supporting translation from Chinese into readable English as well as from English into readable Chinese (Loh, Hung and Kong 1977).

All the versions of CULT mentioned above are in batch mode. One of the major problems of this mode of translation is the problem of inputting Chinese characters. Firstly, the Chinese text was manually encoded into four-digit standard telegraph codes and punched onto cards before being inputted into the computer for translation. This process is enormous, tedious and liable to errors. In order to solve this problem, a research project was carried out at the Chinese University of Hong Kong to try to find a better solution to this problem. As a result, a Chinese key-board was designed.

The newly constructed Chinese key-board makes the design of the present on-line system possible. The use of this key-board for direct input has eliminated many of the errors arising in coding Chinese characters either by means of numerical digits or alphanumerics, thus facilitating the translation process.

The On-line Machine Translation System

The present on-line system is primarily based on the CULT (batch mode) system just mentioned, except that the pre-editing procedures can be replaced by the interactive operations of the Chinese key-board, and also the editing facility

is provided.

The flowchart of the general translation procedures is given in Figure 1, similar to CULT (batch mode) system. Figures 2,3,4,5,6,7,8 & 9 give the detailed flowcharts of (i) Source text preparation, (ii) Input, (iii) Lexical Analysis, (iv) Syntactic & Semantic Analysis, (v) Relative order analysis, (vi) Target equivalence analysis, (vii) Output and (viii) Output refinement respectively.

The editing facilities for on-line up-dating of dictionary records, i.e. correction, insertion, deletion etc. are provided at appropriate points. This enables the operation to interact with the translation system at all times, while translating. Programs for the system are written in Standard FORTRAN and run on the PDP 11/34 computer system.

Computer Configuration

A PDP 11/34 computer system has recently been installed at the Hung On-To Research Centre for Machine Translation, to be used exclusively for the Machine Translation Project. The configuration of the system is given in Figure 10. The Chinese key-board is attached in addition to the standard alphanumeric key-board which is to be used in the implementation of On-line Dual Language Translator in due course. When implemented, the system should be capable of translating Chinese into readable English as well as from English into readable Chinese simultaneously.

The Chinese Key-board

Many Chinese input systems based either on Corner Shapes of the characters, or on Pin Yin (pronunciation), or on a Chinese typewriter key-board with a few thousand keys are commercially available. The widely publicized cylindrical Chinese typewriter key-board which has been developed by the research workers of the Chinese Language Project, Cambridge University, is the latest one to be introduced to the market.

None of these input systems mentioned possesses the most desirable characteristic of being easy to learn and operate. Most, if not all, require extensive training of the operator. For simplicity, the input of Chinese characters to the computer system should follow the natural order of how they are written by hand - the way children are taught to write Chinese character at school.

This kind of stroke-approach, though theoretically sound, is impossible to implement due to the richness of the language and the infinite combination of strokes that go to make up Chinese characters. However, a feasible way is to employ the "radical" approach, in other words, to represent on the key-board the most frequently used radicals of the Chinese characters.

For the past few years, the structures of the Chinese characters have been studied and a new set of modified radicals which total barely over 200 in number, has been selected. A key-board based on this design approach has been constructed and initial test results are satisfactory. The arrangement of the keys on the key-board is usually according to their normal position in the character, that is, a radical, if normally at the top of the character is placed on the upper portion of the key-board and so on.

The main advantage of this approach is that it is simple and natural for any person with a minimal knowledge of written Chinese to operate the key-board with ease.

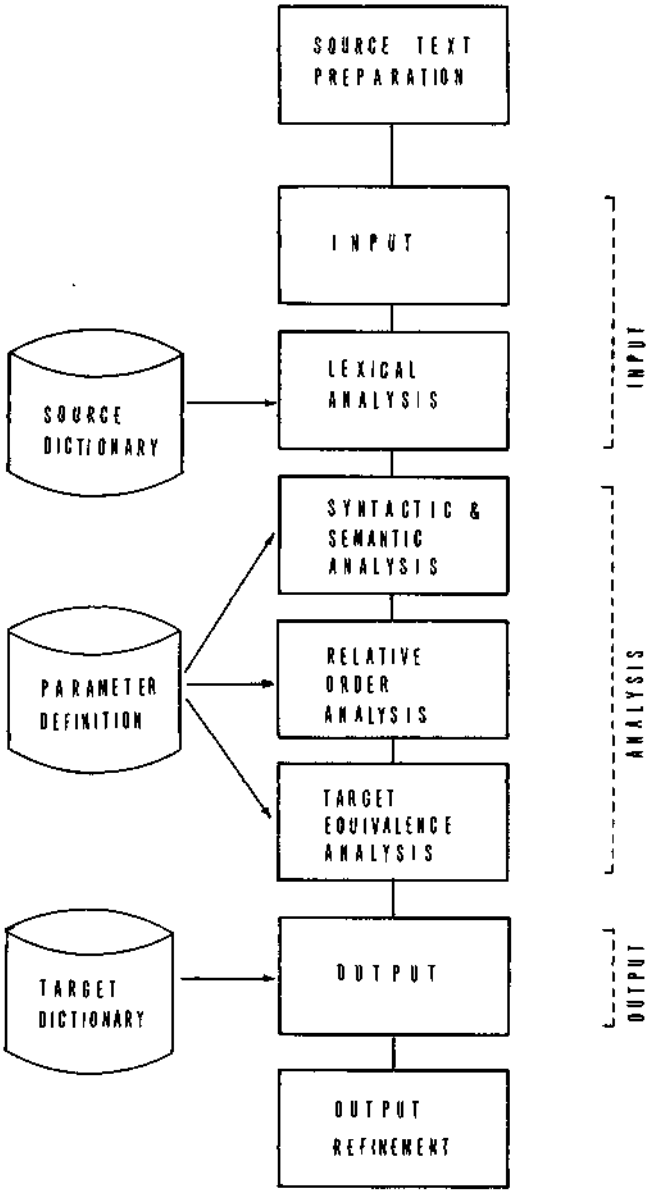


Fig. 1 Translation Procedure

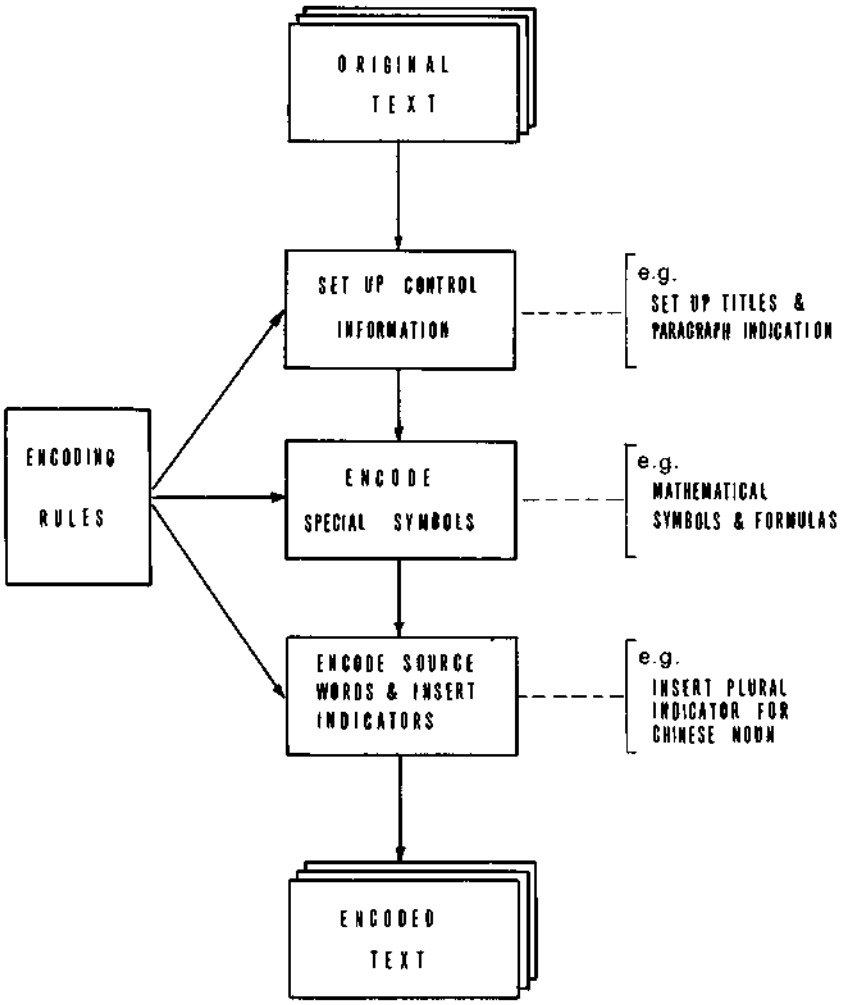


Fig. 2 Source Text Preparation

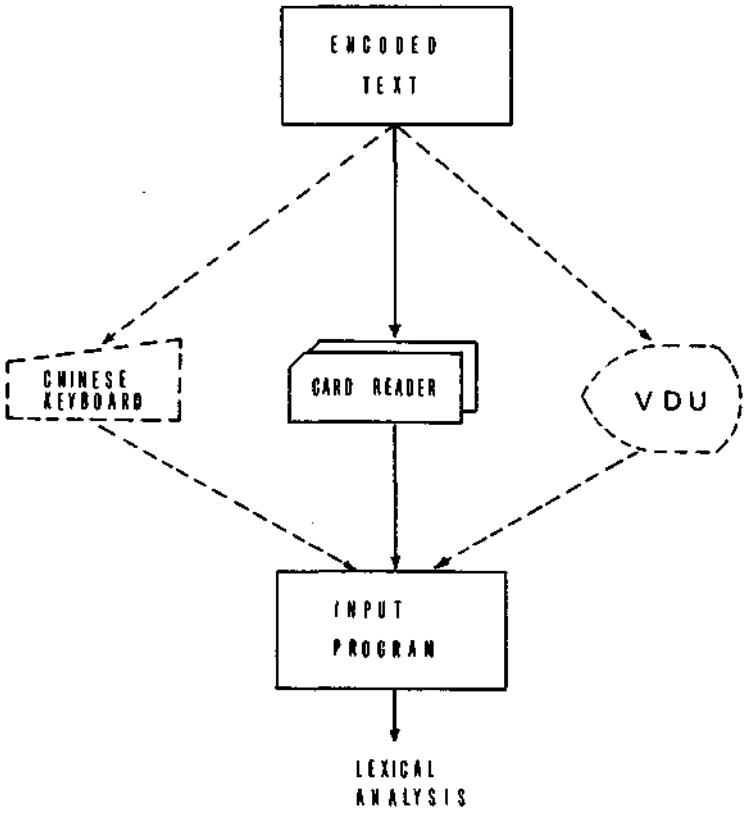

ENCODED
TEXT
CHINESE
KEYBOARD
CARD READER
VDU
INPUT
PROGRAM
LEXICAL
ANALYSIS

Fig. 3 Input

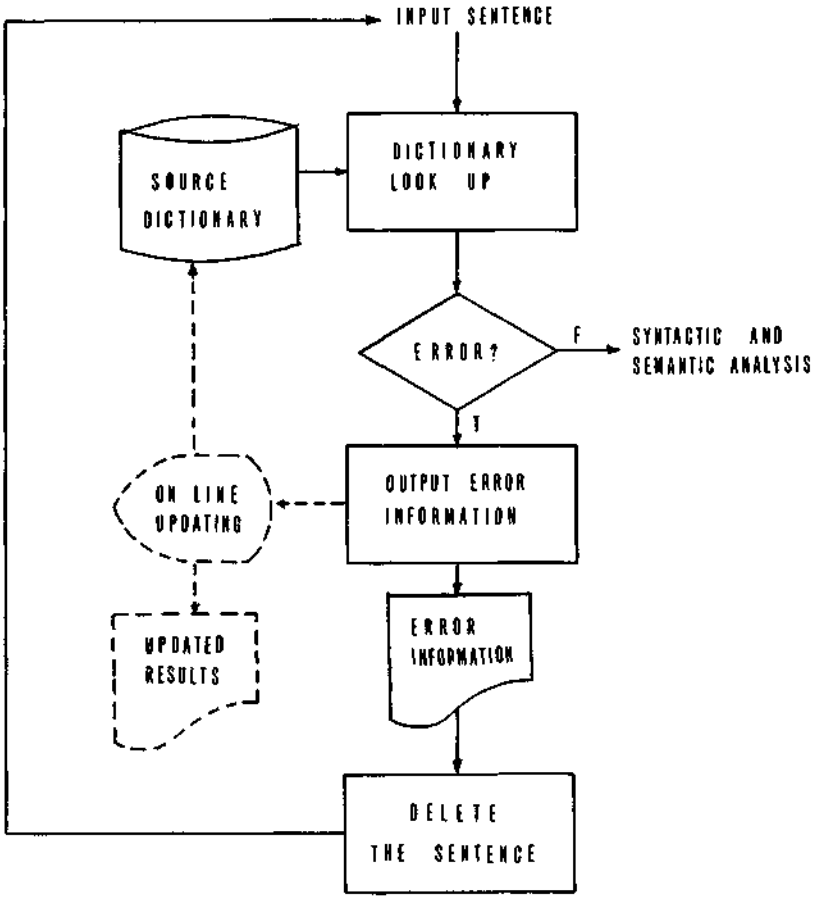


Fig. 4 Lexical Analysis

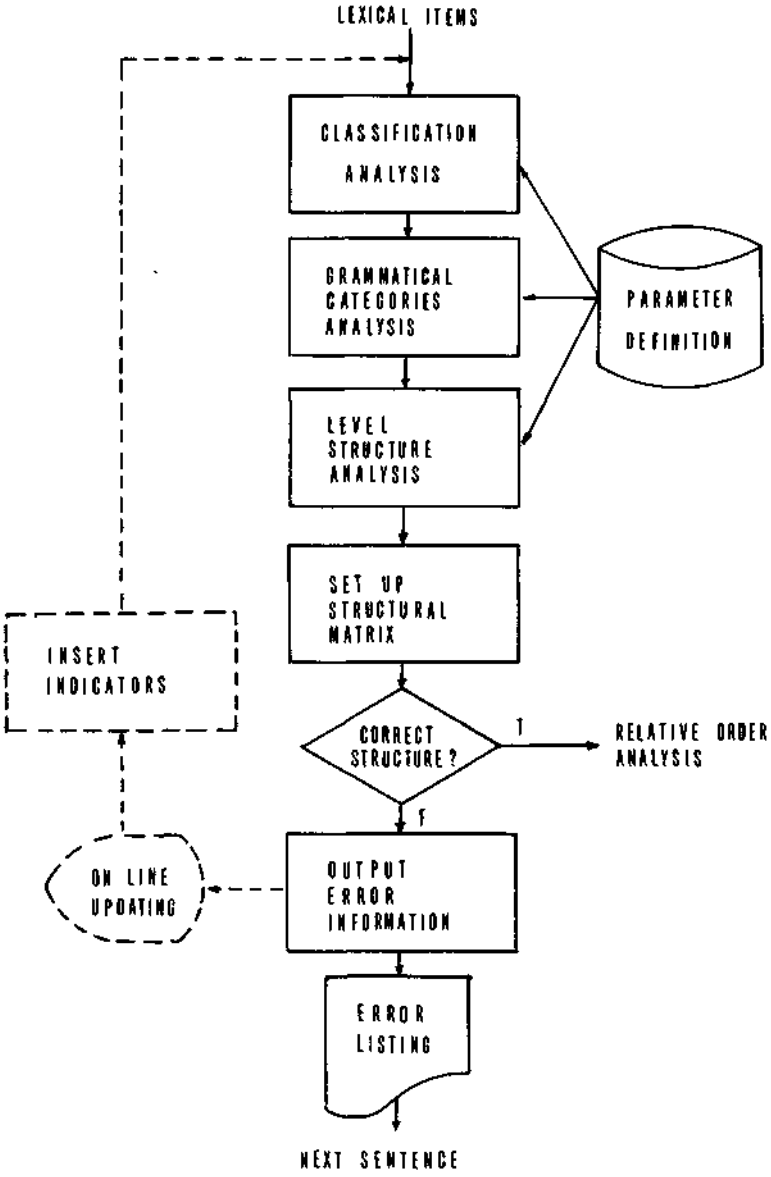


Fig. *5* Syntactic & Semantic Analysis

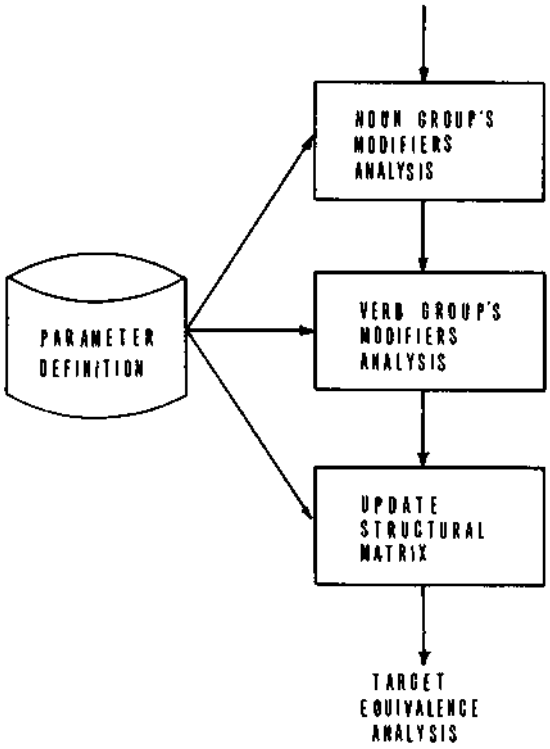


Fig. 6 Relative Order Analysis

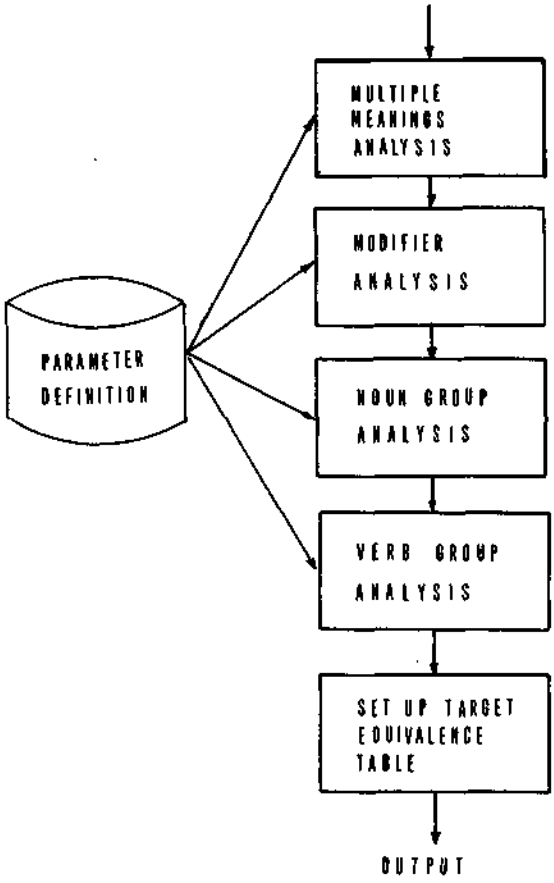


Fig. 7 Target Equivalence Analysis

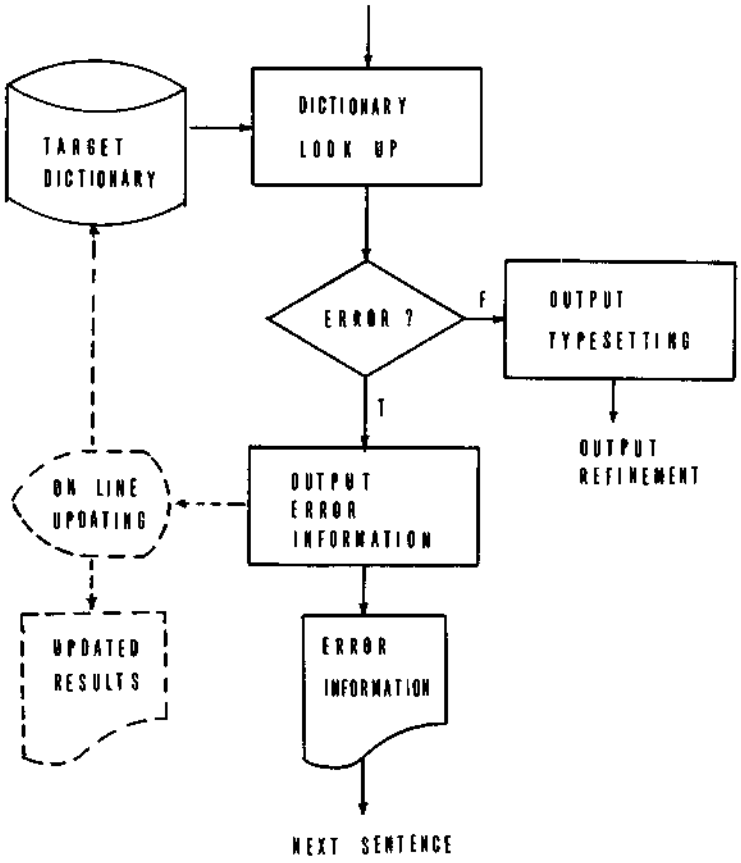

TARGET
DICTIONARY
DICTIONARY
LOOK UP
ERROR ?
F
OUTPUT
TYPESETTING
OUTPUT
REFINEMENT
T
ON LINE
UPDATING
OUTPUT
ERROR
INFORMATION
UPDATED
RESULTS
ERROR
INFORMATION
NEXT SENTENCE

Fig. 8 Output

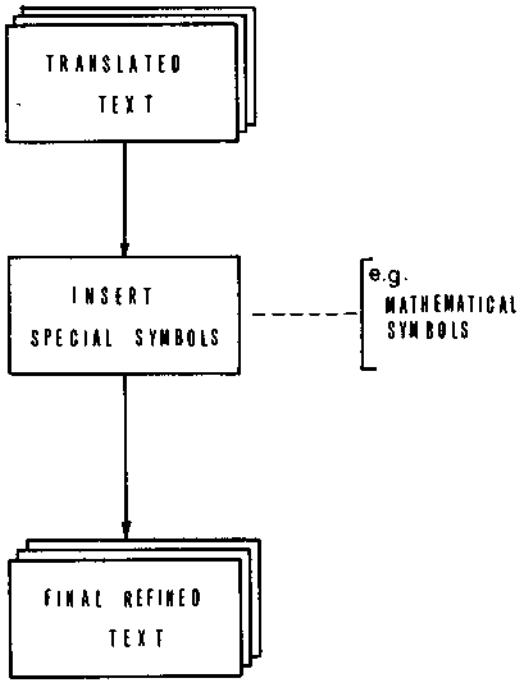


Fig. 9 Output Refinement

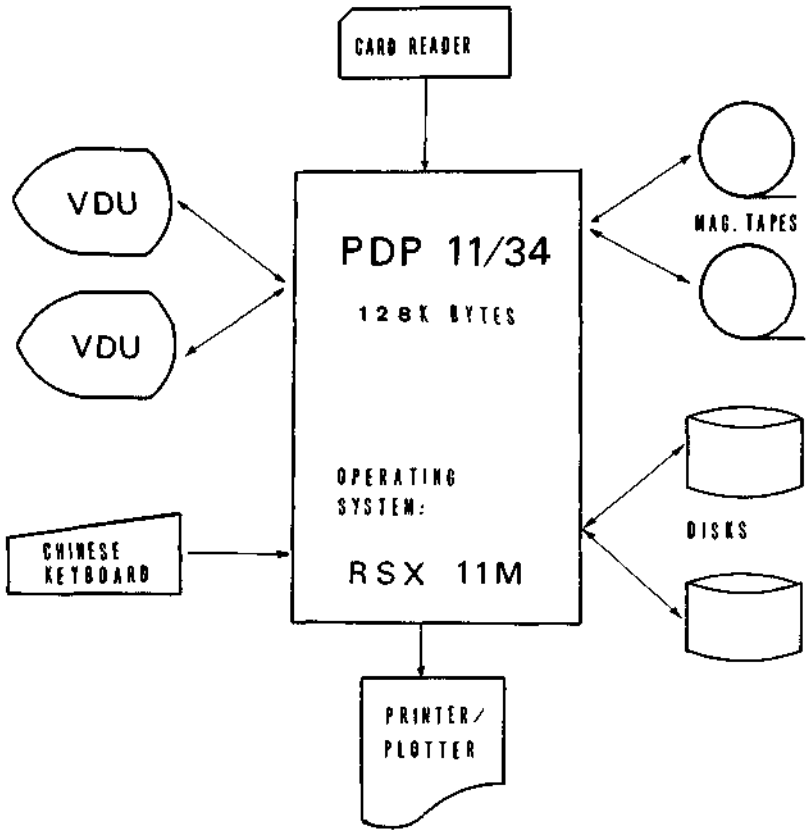
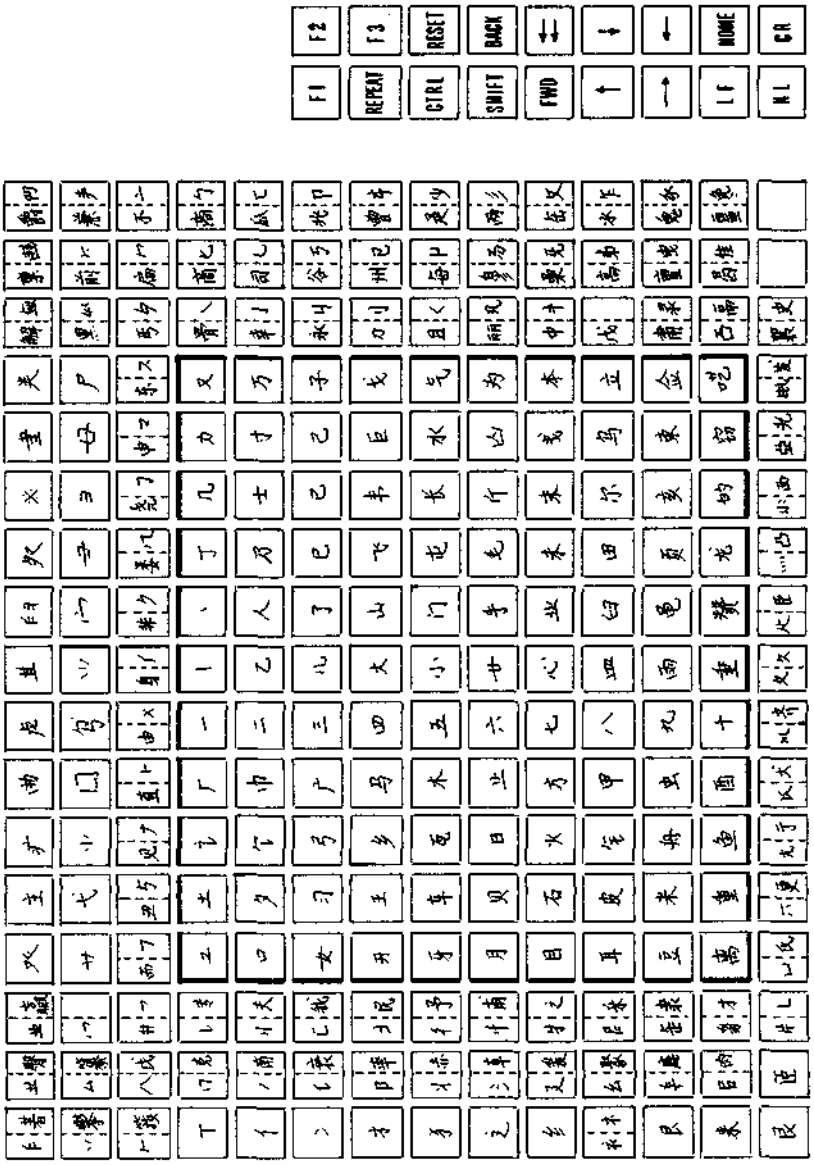


Fig. 10 System Configuration

CHINESE CHARACTER KEYBOARD LAYOUT



Remarks

The implementation of the Chinese key-board has greatly enhanced the throughput, and thus has enhanced the efficiency and performance of the translation system.

Acknowledgements

The authors are grateful to The Asia Foundation and Rockefeller Brothers Fund for initial financial support of the Machine Translation Project and to Hung On-To Memorial Fund for their generous financial assistance in setting up the research centre.

References

Loh, S.C. (1972) Machine translation at the Chinese University of Hong Kong. Proceedings of the CETA (Chinese-English Translation Assistance) Workshop on Chinese Language and Chinese Research Materials, CETA-72-01, 1972

Loh, S.C, Hung, H.S. and Kong, L. (1977) A dual language translator. The Fifth International Symposium on Computers in Literary and Linguistic Research, Birmingham, April 1977.

Loh, S.C. and Kong, L. (1977) Computer translation of Chinese scientific journals. Proceedings of the Third European Congress on Information Systems and Networks "Overcoming the Language Barrier", Luxemburg, 1977.