

DESIGN ISSUE OF ECTST

Huang Jianshuo

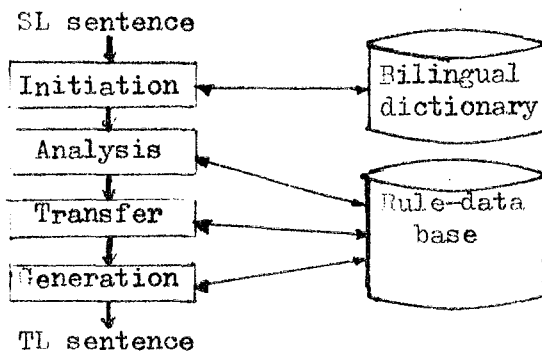
Language Training Center
South China University of Technology
Guangzhou 510641, P.R.C.

ABSTRACT

This note presents an overview of the English-Chinese Translation System for Tourists (ECTST) currently under development at South China University of Technology. A brief description of the bilingual dictionary is given, followed by descriptions of grammar rules representation and the main processes involved in translation.

SYSTEM CONFIGURATION

ECTST consists of a translation program, a bilingual dictionary and a rule-data base. The program body is composed of a sequence of modules, performing translation in four phases:



The initiation phase is concerned with input of the SL sentence, dictionary look-up and morphological processing. The analysis phase involves syntactic and semantic parsing, which are accomplished through linguistic models and case frame. The transfer phase comprises rules for converting the parsed SL sentence into the TL sentence. The

generation phase, applying the contents obtained from the previous process, generates the TL sentence.

ECTST is implemented in PASCAL. The software is separated from the linguistic data, so that any language can be accepted if the grammar rules and vocabulary are given. Likewise, the system program are free to change their components with no undue restrictions that may be imposed by the linguistic components.

DICTIONARY CONTENTS

The dictionary is bilingual: it contains morphological, syntactic and semantic information needed for the analysis and generation of a sentence. This includes information about the category of words, their semantic features and case frames. Entries in the lexicon are of eight types, and their semantic features are based on graded concepts. The case frame provides a means to find out the logical relationship between components in the sentence.

At present, the dictionary contains a limited set of lexical entries, which are grouped into models with the initial letter as the index for access.

GRAMMAR RULES

In ECTST, rules are represented in the form of meta-model. It may contain one or more data items as shown below:

A [x](a) B [y](b) C [z](c)

where A,B,C represent information on category of the word, while x,y,z and a,b,c are parameters indicating multiple pieces of information and tree structure separately. In many cases, semantic features or information on grammatical gender, number or case can be used as parameters.

Generally, hundreds or even thousands of grammar rules are necessary for language analysis and transformation. It is impractical to check whether all these rules can be applied to all intermediate processings. To make the processing efficient, rules are grouped into packets, each dealing with specific corresponding linguistic phenomena ranging from a phrase, a clause to a sentence.

ANALYSIS

The work of analysis is performed by the parse module. It constitutes a unified syntactic and semantic analysis of the input sentence, and its output serves as an input to transfer. The parsing program interprets the input sentence according to the available linguistic models. These come in several varieties, such as meta-models, case frames etc..

The parse module consists of a number of routines, each responsible for a special work. One routine for example performs morphological analysis of words; another, for syntactic and semantic analysis of phrases; still another, for application of case frames, and so on. Analysis are carried out from the surface structure through the deep structure. A deep structure may comprise roles of six essential types: SUBject, OBJect, INVolvement, SITuation, STAte and CAU-sation. They make up a basic frame with the PREDicator as shown in Fig. 1.

In application, a role may come into its subcategories whenever necessary. SITuation, for example, may fall into

TIME, LOCation, DEGREE or CONDition; INVolvement into RECipient, PARTICipant or GOAL; and CAU-sation into BASIS, PURPOSE or RESULT and so on.



Fig. 1

In analysis, the deep structure is obtained via preferential weightings calculated from grammatical data and frame labels. The following sentences for example would result in two different internal structures even though their surface structures are similar.

1. We found John very well.
2. We know John very well.

In the case frame, their deep structures would be:

1. found (SUB(1),OBJ(2),COM(3))
2. know (SUB(1),OBJ(2),STY(3))

When an internal structure results, it serves as the basis for transfer.

TRANSFER AND GENERATION

In the transfer stage, transformation is accomplished in two steps: first the internal structure of the SL sentence is generated into a tree, with nodes that indicate their proper semantic and grammatical order in the TL sentence; then lexicon rules are invoked to transfer the SL entry, on the basis of its context, into the TL entry. The advantage of such a procedure are obvious. It makes the rules more flexible, especially those coping with sentences with similar internal structures. Moreover, rules can be added, modified or changed as needed with no resultant effect on one another.

When a target tree is built up, it is then scanned by way of left recursion

and with words appended to its terminal nodes as required. Finally, a character string is obtained; its output is a sentence in TL.

CONCLUSION

Efforts made over the past decade have achieved considerable progress in machine translation. First syntactic parsing was pursued, then semantic and context analysis was advocated. ECTST has profited from both theories.

Natural language is essentially a discrete information system. It often allows multiple syntactic interpretations. To minimize the possibility of multiple interpretation, we introduce a minimum amount of semantic information and adopt the case frame. This provides flexible facility in syntactic analysis and helps to distinguish structural ambiguity. In this approach, the translation accuracy is up to 90% or more, which can be raised if corresponding information and/or rules are modified and supplemented.

REFERENCES

1. Pan Yu and Yuan Y. Sung (87). "Automatic English-Chinese Translator," Proceeding of 1987 International Conference on Chinese Information Processing, pp. 52-61
2. Jonathan Slocum Siemens Corporation and Winfield S. Bennett (82). "The LRC Machine Translation System: An Application of State-of-the-Art Text and Natural Language Processing Techniques to the Translation of Technical Manuals," Working Paper LRC-82-1 July 1982
3. Mei-Du Li and Gary G. Erickson. "Modular Dictionary Design For Chinese-To-English Machine Translation," GARJAK Research Inc. 990 Highland Drive, Solana Beach, California 92075-2472, pp. 249-252
4. Huang Jianshuo (87). "Language Features and Semantic Information in MT," Proceedings of ICCIP'87, pp. 110-116
5. Makoto NAGAO (82). "A Survey of Natural Language Processing and Machine Translation in Japan," Computer Science & Technology, pp. 64-70 T. Kilagawa(ed.) CHMSHA, LTD. & North-Holland Pub. Co.
6. Makoto NAGAO (86). "Current Status and Future Trends in Machine Translation," Future Generations Computer System 2 (1986) pp.77-82. Elsevier Science Publishers B.V. (North Holland)