

Machine Translation Based on Translation Rules for Processing Natural Language

Li Zhou

Department of Foreign Languages and Tourism, Guilin Normal College, Guilin 541001, China

Zhouzi@163.com

Keywords: natural language processing, machine translation system, hybrid, example-pattern-based, rule-based

Abstract. This paper presents a hybrid approach, which integrates an example-pattern-based method and a rule-based method, to the design and implementation of an English-Chinese machine translation system. It focuses discussion on language model, knowledge base, design ideas and implementation strategies. Our system has been tested based on requirement details listed in the Outlines for Automatic Evaluation of Machine Translation constituted by National Hi-Tech Project 863, and compared with the Huajian system. Experiment results indicate that our system has high translation speed and accuracy.

Introduction

Machine Translation refers to a computer from one natural language to another translation of Natural Language, means a natural language with equivalent transformation[1]. Machine translation technology since in the 2040 of the 21st century - received worldwide attention, especially as the degree of information to enhance the social life and the extensive international exchanges and cooperation, and people have a higher speed of access to information and greater demand for cross-language information, low efficiency of human translation is less and less to the development of the information society[2]. Therefore, the machine translation technology is emerging as hot spots of information science. Describes the language of system - related knowledge base, describe the system design of the main module and system performance test results are given at the end[3].

Whole Structure of System

The System mainly includes the translation of knowledge base systems, subsystems, knowledge management and user interface. Systems Architecture as shown in figure 1. One of the main body of knowledge base system and subsystem is a system of translation, respectively, based on models of linguistic knowledge and the translation model. Rule based system model analysis method as the main source language, into a logical analysis, and generates two phase converter, the method using a source of knowledge for the various integrated dictionaries and rule base. We will expand the database instance model library as an example, that the library structure similar instance of an abstract model of the sentence containing the arguments, and stored in the library. Translation, in addition to the input sentence and library instances of the source language similarity matching portion, will also be part of the source language input sentence and repository pattern similarity match.

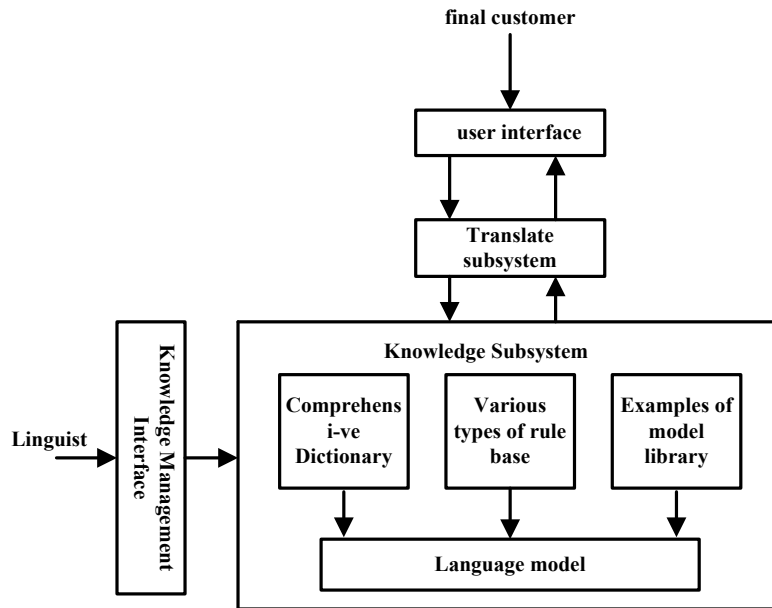


Fig.1 Whole Structure of System

Method of converting generation system based on rules as the main body, to embed method of case - based model parsing phase. System priority of the output methods of case - based model of translation results. When the instance is not in the pattern library instance similar to enter the source language sentences and sentence mode, the system output is based on the rules of translation results. Combination of two ways as shown in figure 2.

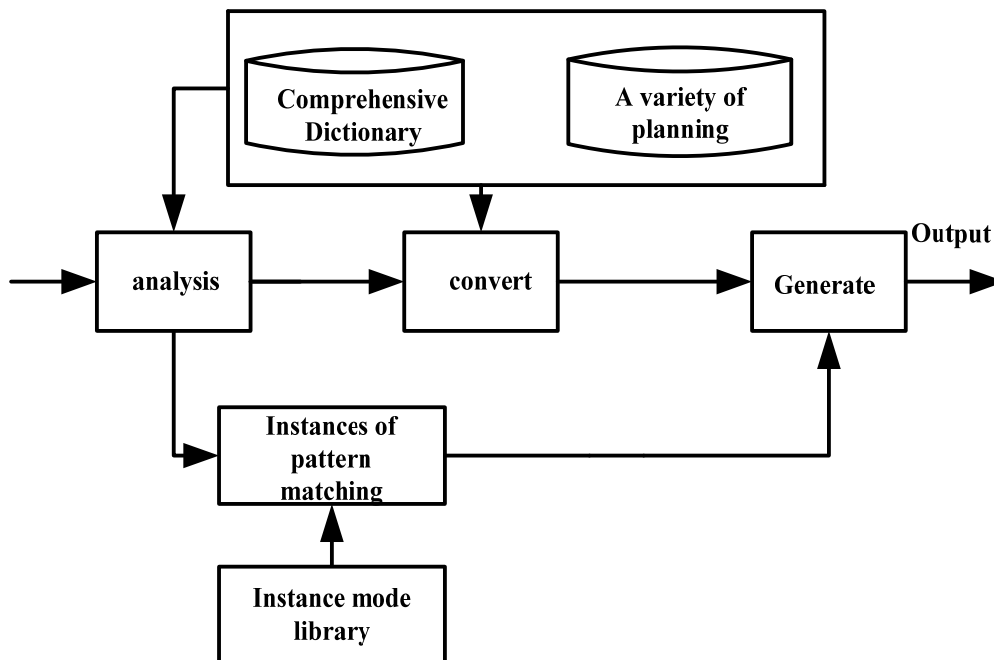


Fig.2 Combination of two methods of translation diagram

Translate subsystem

Translation of system subsystem consists of 8 interrelated functional modules such as lexical analysis, the logical relationships as shown in figure 3. Of these, lexical analysis, part-of-speech tagging, shallow parsing module 3 is the analysis of the source language, is the common basis of two methods of translation module. Instances of pattern matching module is based on the concrete

embodiment of the method instance mode. Sentence transform to the source language and deep Syntactic Analysis is further analysis, preparing to convert. Sentence structure Conversion module source to the target language conversion. Phrase of the target building modules are two methods belong, that is, the output target translations.

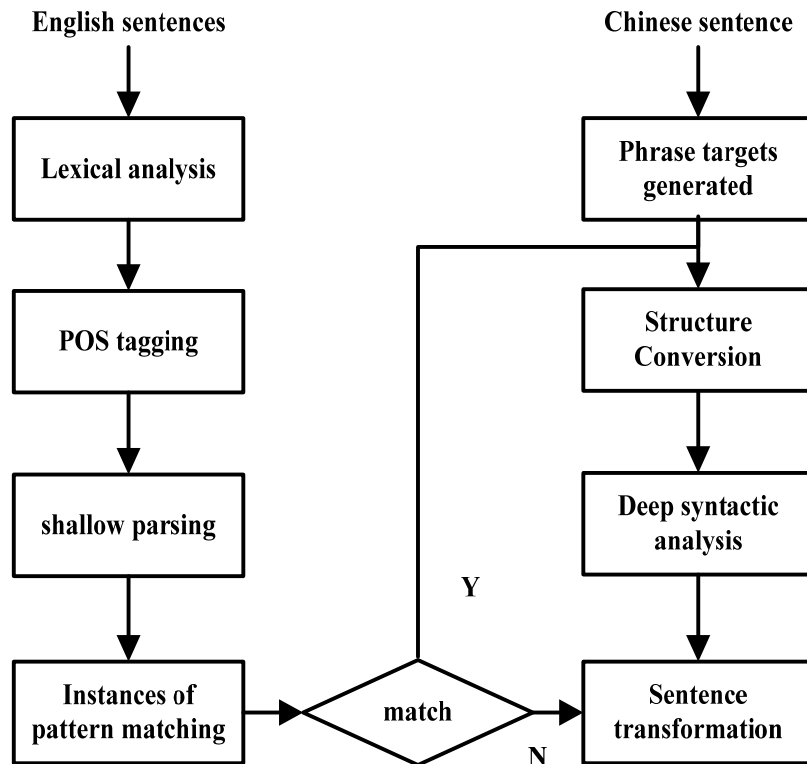


Fig.3 System sketch of the translation process

Part of speech tagging is based on a certain context, to enter the automatic annotation of speech of a word in the sentence. The system uses part-of-speech tagging method based on combination of rules and statistics. Part-of-speech tagging using POS rule based on the rule bound to the input array of speech, from the perspective of the entire sentence, remove the invalid part and morphological information, forming part of speech sequence. Based on statistical part-of-speech tagging using model parameters extracted from the corpus in advance, as well as contextual information, part of speech and further eliminate such ambiguity.

Lexical analysis and POS tagging

Lexical analysis is based on rules with linguistic rules as source of knowledge, transforming the real language text to process a sequence of speech. Lexical Analyzer includes pre - processing system, Morphological analysis, not a few features such as word processing and analysis of the phrase sub module[4],mainly use the comprehensive dictionary of knowledge source system, derived vocabulary, morphology rules, rules of word formation rules and groups of words, derived vocabulary of which are the result of the lexical analysis of self - learning. Structural Model of lexical analyzer as shown in figure 4.

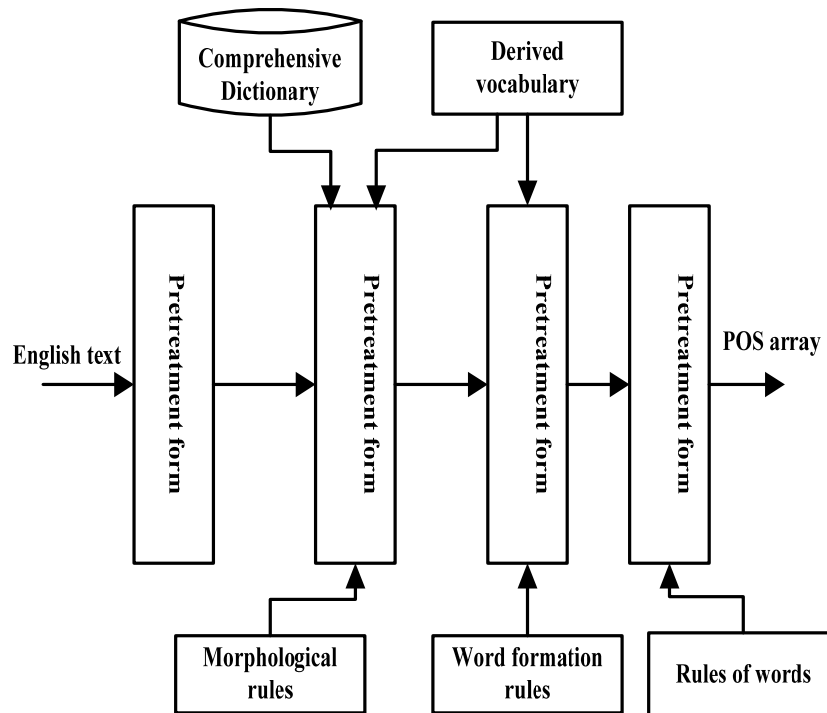


Fig.4 Lexical analyzer structural model

Phrase structure and meaning of the phrase objectives generated key choices, such as Quantifiers add. Phrase structure using Chinese phrase structure adjustment of knowledge in the phrase word order. Selected primarily on the basis of lexical context of meaning, using the comprehensive settlement of the polysemous words the dictionary translation selection issue. Adding quantifiers are based on the structure of the phrases and phrases of the word semantic properties determine what do you want to add, and add quantifiers.

Summary

The combination method based on rules, design and implementation of an English - Chinese machine translation system based on hybrid strategy. Experimental results show that the system has asked the correct rate is high and the characteristics of the translation speed. However, there is a lack of linguistic knowledge systems, inadequate poor capacity for handling the complex sentence, you still have to continue in the relevant areas of research study and practice.

References

- [1] B.M. Wu: *Matlink English - Chinese machine translation system* (Ph.D., Zhengzhou : Information Engineering University of PLA, China 2010), p.34.
- [2] T.J. Zhao: *Principle of machine translation* (Harbin Institute of Technology Press, China 2009).
- [3] Y.H. Guo: *Computer applications*, Vol. 24(2008) No.3, p.71.
- [4] Z.P. Du: *Journal of Information Engineering University*, Vol. 4(2010) No.3, p.89.