

DESIGN AND IMPLEMENTATION OF A CHINESE-TO-ENGLISH SPOKEN LANGUAGE TRANSLATION SYSTEM¹

Chengqing ZONG, Taiyi HUANG and Bo XU

National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing
{cqzong, huang, xubo}@nlpr.ia.ac.cn

ABSTRACT

In this paper, we describe the design and implementation of a Chinese-to-English spoken language translation system. The system employs the multiple translation engines, and it consists of a template-based translator, a semantic parsing based translator (SPBT) and a statistic based translator (SBT) as well. SPBT uses the interchange format (IF) to represent the understanding results of input utterance. The target language generator generates the translation results according to IF. The dialog knowledge manager is designed to record the dialog history and help the Chinese parser to find the topics of the analyzing utterance. Now the system is under construction, and it is restricted in the domain of hotel reservation. Some preliminary experimental results are reported in the paper.

1. INTRODUCTION

Although the spoken language translation (SLT) systems have broken many barriers in the past over decade^[1,2,3], the current SLT system still suffers from the performance limitation even if the domain is restricted. The authors think the translation strategy and the approach for parsing spoken language are two key factors to affect the performance of SLT system.

As representatives of SLT systems at present, Verbmobil^[2,4], JANUS-III^[1,5] and ATR-MATRIX^[3] represent the main translation approaches. In ATR-MATRIX system, only one translation engine (*Example-based*) is employed. In Verbmobil system, although multiple parsers cooperate with each other, the translation approach is still unitary (*interlingua-based*). The authors believe the unitary translation approach is often limited to cope with the utterances with much ambiguity, and the system's flexibility to parse various types of utterances is also limited. JANUS-III system employs two alternative translation modules, which are direct translator and interlingua-based translator, but it seems unreasonable that the parsing process is not concerned with dialog history knowledge. We think the dialog knowledge manager is an indispensable module in a spoken language translation system.

On the other hand, the parsing approach is always the sticking point of a spoken language translation system. As you know that all parsing approaches at present are of their own faults. So, how

to combine the different approaches and how to remedy their faults are still remained to further research.

In this paper, a Chinese-to-English spoken language translation system is presented, which is restricted in the domain of hotel reservation, and limits the vocabulary to about 12000 words in Chinese and about 2000 words in English. The system employs multiple translators, and the Chinese parser works under the guidance of dialog knowledge manager. Remainder of the paper will describe the system in detail. *Section 2* gives the overview of the system. *Section 3* shows some recent experimental results, and *Section 4* gives the concluding remarks.

2. THE SYSTEM OVERVIEW

The system consists of three component parts: (1) the Chinese speech recognizer (CSR), (2) the Chinese-to-English translation (CET) module and (3) the English speech synthesizer (ESS). CET module includes five sub-modules, and they are: (a) the preprocessor; (b) the template-based translator (TBT); (c) the semantic parsing based translator (SPBT); (d) the statistic based translator (SBT); and (e) the dialog knowledge manager (DKM). SPBT uses interchange format (IF) to represent the understanding results of input utterances. The target language generator (TLG) generates the translation results according to IF. CSR outputs an N-best list of Chinese text string hypotheses. CET module gives the English text corresponding to each input utterance. ESS converts the translation results into speech. The component diagram of the system is shown in Figure-1.

The five sub-modules of CET module are described in detail as follows.

2.1. Preprocessor

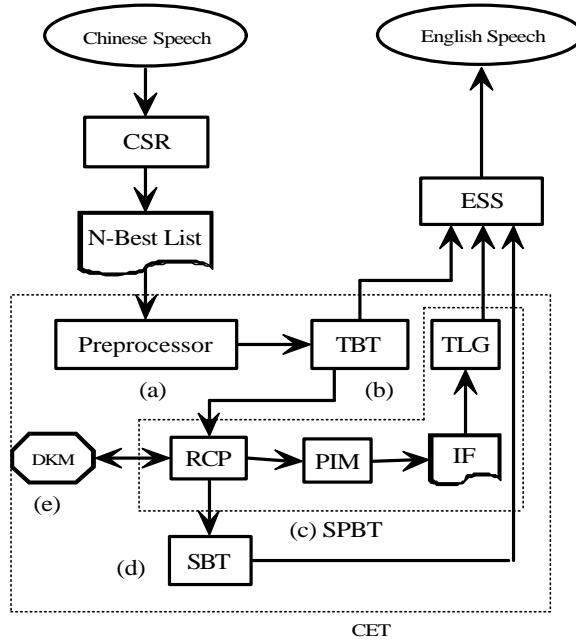
The preprocessor is designed to process the utterances before the template-based translator. In our system, the preprocessor completes the following three functions mainly:

- To delete all repeated words except some special adverbs. In the same Chinese sentence, except some special adverbs like ‘非常 (very)’, ‘特别 (very, in particular)’ etc., the same words are generally not

¹ The research work described in this paper is supported by the National Key Fundamental Research Program (the 973 Program) of China under grant G1998030504, the National Natural Science Foundation of China under grant 69835030 and also the National 863 Hi-Tech Program under grant 863-306-ZT03-02-2.

repeated continually. Even if some special words are repeated continually, the number of times of repeated words is usual not big than two. So, we designed a special base to store the special words. If a word is recorded in the base, the word may be continually repeated two times. Otherwise, any words are not allowed to repeat continually. According to the principle, all extra repeated words in the input utterances will be deleted by the preprocessor.

- To recognize and analyze the numerals and numeral phrases in the input utterances. The numeral phrases are signed as QP. And the Chinese numerals are all translated into Arabic numerals.
- To recognize and understand the time words or time phrases in the input utterances. The time words are reduced into the phrase TP and translated into English expression^[6].



- RCP: Robust Chinese Parser
- PIM: Parser-to-IF Mapper

Figure-1. Architecture of the System

2.2. The Template-based Translator

The template-based translator is the first translation engine in our system. In the system we proposed an improved template-based approach to spoken language translation. The template-based translator uses flexible expression format to describe the template condition. The template is designed as:

$$C_1 \ C_2 \ \dots C_n \Rightarrow T_l \quad \dots \text{(1)}$$

Where, n is an integer and $n \geq 1$. C_i ($n \geq i \geq 1$) is a component which expresses a condition that the input utterance of source language has to meet. T_l is the output result corresponding to the input. The formula (1) means if an input utterance of source language meets the conditions $C_1 \ C_2 \ \dots C_n$, the input utterance will be translated into the target language expression T_l . In our system the source language is the Chinese language, and the target language is English.

The condition of a template may consist of keywords, parts-of-speech and also semantic features, so the input may be matched with a template from shallow level to deep level. In the condition of a template, the distance between two fixed keywords is stretchable, thus some needless words or non-keywords in the input utterances may be skipped in matching operation. And also the translation results of the same template are alterable. The proper results are finally generated according to the specific context. That is, the relation between a template and translated utterance is one-to- n (where, n is an integer and $n \geq 1$). At present, the correct rate of translation results of the template-based translator is 51.4%. For detail of the template-based translator, please refer to [7].

2.3. The SPBT Translator

SPBT is the second translation engine in our system, which includes a robust Chinese parser (RCP), a parser-to-IF mapper (PIM) and a target language generator (TLG).

2.3.1. The RCP Parser

The RCP is a robust parser for Chinese spoken language understanding with semantic grammar. The RCP was inspired by partial parser^[8,9] and chunk parser^[10], and is a bottom-up parser of head word-driven probability context-free grammars. The parsing rule is expressed as the following format^[11]:

$$A_j(s_{j1}, s_{j2}) \dots A_k(s_{k1}, \dots, s_{km}) \rightarrow XP, !H \mid P \dots \text{(2)}$$

Where, k , m_j , and m_k are all integers and $k \geq 1$, $m_j \geq 1$, $m_k \geq 1$. A_i ($k \geq i \geq 1$) is a part-of-speech, s_i ($\geq i \geq 0$) is a semantic feature. XP is the new phrase name. H is the part-of-speech of the head word in the phrase, and H is one of A_i . P is the probability or frequency of the rule. See the following examples:

$$\begin{aligned} AP + NP &\rightarrow NP, !NP|64 \\ NP1(nsh)+NP2(nsp)+'人' &\rightarrow CS, !VP|12 \end{aligned}$$

In the construction of parsing tree, when a new node is generated, a related semantic rule will be called to annotate the new node. That means, each semantic action is performed accompanying with the related syntactic reduction. Thus, each non-terminal node is annotated by semantic concept. Three examples of semantic rules are shown in the following:

$$\begin{aligned} time-word &\rightarrow \{TIME\} \\ \{TIME\} + \{TIME\} &\rightarrow \{TIME\} \\ \{PLACE\} + \{TIME\} &\rightarrow \{PLACE, TIME\} \end{aligned}$$

Where, *time-word* is a concrete Chinese time word or phrase. $\{PLACE\}$ and $\{TIME\}$ represent the ‘place concept’ and ‘time concept’ respectively. $\{PLACE, TIME\}$ means ‘place concept’ and ‘time concept’ coexist.

2.3.2. PIM, IF and TLG

The results of RCP are represented by parsing trees, in which all non-terminal nodes are annotated by semantic concepts. PIM converts the parsing results into IF.

IF represents the semantic meanings of input. The meanings of each input are represented by four components^[12]:

```

IF ::= speaker: speech-act: topic: argument
speaker ::= agent | customer
speech-act ::= give-information | request-information | ...
topic ::= (concept = attribute) ^
argument ::= (concept = attribute) | *

```

Where, both *topic* and *argument* are made up of attribute-value pairs. The attribute may be any concept defined in the specific domain. The value may be an atomic symbol or recursively an attribute-value pair. The symbol ‘^’ indicates that the expression may appear zero to one times. While symbol ‘*’ means the expression may appear zero to any times. And the attribute value pairs are order free. See the following example:

```

agent: give-information: (available = (room = (room-type =
double))): (price = (quality = 200 & 240, currency =
dollar))

```

The expression indicates the corresponding input is uttered by agent, and it means ‘ we have two types of double rooms available, whose prices are respectively 200\$ and 240\$ per-night.’ .

The TLG is an English language generator, which generates the English text according to the IF frame. In TLG, the template-based method and the technique of rule-based text planning are employed.

2.4. The SBT Translator

SBT is the third translation engine in the system. In our system, the STB translator employs the common statistical opinions. The translation task may be described as follows. Given a source (‘Chinese’ here) string $c_1^m = c_1 \cdots c_j \cdots c_m$, the target (‘English’ here) string e_1^l , with the highest probability, is chosen among all possible target strings $e_1^l = e_1 \cdots e_i \cdots e_l$ by use of Bayes’ decision rule:

$$\begin{aligned} \hat{e}_1^l &= \arg \max_{e_1^l} \{\Pr(e_1^l | c_1^m)\} \\ &= \arg \max_{e_1^l} \{\Pr(e_1^l) \cdot \Pr(c_1^m | e_1^l)\} \quad (3) \end{aligned}$$

$\Pr(e_1^l)$ is the probability of the language model produced by the target language. $\Pr(c_1^m | e_1^l)$ is the probability of the string translation model from the target language to the source language. Please see [13] for detail of the translator.

2.5. The Dialog Knowledge Manager

The dialog knowledge manager in our system is designed to record and manage the dialog history, which is mainly consisted of the topics and utterance type (acknowledgement, statement or query etc.) etc. The dialog topic is of heredity, and the topics of current utterance will be inherited to the next utterance until they are

changed or the dialog is finished. The dialog knowledge manager in our system works as a finite state automata (FSA):

$$FSA = (S, t, d, s_0, F)$$

Where, S is a set of all possible states that are represented mainly by the topics in the domain; t is the topic and type of an input; d is a set of conversion rules, according to which the state is changed; s_0 is the initial state and it is general null; F is the end state and is usually indicated by the input of Chinese word ‘再见(Good-bye)’ .

3. Experimental Results

Although the whole system is still under construction now, some components of the system have been realized, which includes CSR, TBT, SBT and also the preprocessor. The preliminary test results are presented in this section.

The test corpus was collected from the domain of hotel reservation, but it is not included in the training corpus. Table-1 gives the survey of the test corpus.

TOTAL NUM. OF UTTER.	NUM. OF CHINESE WORDS	NUM. OF CHINESE CHAR.	WORD NUM./UTTER.
70	441	662	6.3

Table-1 The Test Corpus

The following Table-2 shows the performance of our CSR.

CASES	WORD		SENTENCE	
	Top=1	Top=10	Top=1	Top=10
CORRECT RATE (%)	90	95	70	85

Table-2 Performance of the CSR

Under the accuracy of top ten candidate results of the CSR, and with 106 templates and 450 parsing rules, the correct rate of TBT reaches over 50%. The test results of the TBT are listed in Table-3.

Results	Correct	Part of Result is Correct	No Result
Num. of Utter.	36	7	27
Ratio (%)	51.4	10.0	38.6

Table-3. Translation Results of TBT with Input of Top Ten Results of Speech Recognition

The 106 templates are summarized from 1514 utterances, and the 450 parsing rules are abstracted from 94 dialogs (3013 utterances)^[7].

After training by use of 3009 Chinese-English sentences, the SBT translator gets 21.6% fair translation results, 43.1% acceptable translated results and 35.3% wrong results^[13].

4. CONCLUSION

The preliminary experimental results of our Chinese-to-English spoken language translation system have made us confident of developing the robust and effective SLT system. However, the current research work involves many hard problems. We have been working on the following problems:

- ◊ Improving the coverage of our grammars and increasing the description ability of semantic grammar;
- ◊ Enriching the semantic expression form of the interchange format;
- ◊ To expand further the functions of dialog knowledge manager, and to realize the interaction between speaker and system to help the parsing process.
- ◊ To design the effective function to judge whether the results of RCP are correct.

5. REFERENCES

- [1] Alon Lavie, Lori Levin et. al. The JANUS-III Translation System: Speech-to-Speech Translation in Multiple Domains, *Report in C-STAR Workshop*, Schwetzingen, Germany, 24 Sept. 1999.
- [2] Wolfgang Wahlster, Concurrent Shallow and Deep Processing in the Verbmobil System, *Report in C-STAR Workshop*, Schwetzingen, Germany, 24 Sept. 1999.
- [3] Benjamin Reaves, Atsushi Nishino, Akio Yokoo. ATR-MATRIX: Implementation of a Speech Translation System. *Report in C-STAR Workshop*, Schwetzingen, Germany, 24 Sept. 1999.
- [4] Thomas Bub, Wolfgang Wahlster, Alex Waibel. Verbmobil: The Combination of Deep and Shallow Processing for Spontaneous Speech Translation. In *Proc. ICASSP' 97*, vol. I, pp.71-74.
- [5] Alon Lavie, Alex Waibel and Lori Levin et. al. JANUS-III: Speech-to-speech Translation in Multiple Languages. In *Proc. ICASSP' 97*, vol. I, pp.99-102.
- [6] Chengqing Zong, Taiyi Huang and Bo Xu, Approach to Recognition and Understanding of the Time Constituents in the Spoken Chinese Language Translation. To appear in *Proceedings of the Third International Conference on Multimodal Interfaces (ICMI' 2000)*. Springer-Verlag Press, Oct. 2000.
- [7] Chengqing Zong, Taiyi Huang and Bo Xu, An Improved Template-Based Approach to Spoken Language Translation. To appear in Proc. *ICSLP2000*. Beijing.
- [8] Nam-Yong Han, Un-Chen Choi and Youngjik Lee. An Implementation of a Partial Parser in the Spoken Language Translator. In *Proc. ICASSP' 98*, vol. I, pp.205-208.
- [9] Steven Abney. Partial Parsing Via Finite-State Cascades. In *Proc. ESSLLI' 96 Robust Parsing Workshop*. From <http://www.sfs.nphil.uni-tuebingen.de/~abney/>
- [10] Zhou Qiang. The Chunk Parsing Algorithm for Chinese Language (in Chinese). In *Proc. Joint Symposium of Computational Linguistics*, November, 1999, pp.242-247.
- [11] Chengqing Zong. Research on Strategies for Spoken Language Translation (in Chinese). *Report for Finishing Post-doctoral Program*. Institute of Automation, CAS. April 2000.
- [12] Hua Wu, Taiyi Huang, Chengqing Zong and Bo Xu, Chinese Generation in a Spoken Dialogue Translation System. In Proc. *COLING' 2000*, Germany. Vol. II, pp. 1141-1145.
- [13] CHENG Wei and XU Bo, Statistical Approach to Chinese-English Spoken-language Translation in Hotel Reservation Domain. To Appear in Proc. *ISCSLP2000*. Oct. 2000, China.