

# An INTERLINGUA FOR DIALOGUE TRANSLATION

Hua WU, Taiyi HUANG, Bo XU

National Laboratory of Pattern Recognition, Institute of Automation  
Chinese Academy of Sciences, Beijing  
E-mail: { wh, huang, xubo } @nlpr.ia.ac.cn

## ABSTRACT<sup>1</sup>

An interchange format (IF) suitable for spoken language translation is introduced in this paper. It is a semantic representation of languages and used as a kind of interlingua among different languages. The most obvious characteristics of the semantic representation are its independence of peculiarities of any language and its underspecification. The whole semantic representation has up to four components: speaker tag, speech act, topic and arguments. The development of the interchange format is guided by the corpus of our hotel reservation domain. And the IF has been applied to two languages: Chinese and English. This paper will also discuss the role of the interchange format in our spoken language translation system.

## 1. INTRODUCTION

Spoken dialogue translation is more difficult than written language translation because there are more idioms and more flexible expressions in spoken languages. The same meaning can be expressed very differently in different languages. And some of them can not be translated literally. For example, if the English sentence "Can I help you?" is translated into Chinese literally, the result will be strange. This sentence can be used in many situations such as in stores and in banks. But in Chinese, it can be translated differently according to the different situation. For example, the seller in stores often says: "您要点什么?", which is equal to "Can I help you?" in English.

In dialogues, the speaker always expresses his intent when he utters a sentence [1]. If we catch this information, we can translate the utterance better in target languages. We call this information as speech act. If we use the speech act "introduce-help" to express the sentence "Can I help you?", we can translate this sentence very well according to the speech act and domain knowledge.

This paper describes the interchange format used in our hotel reservation domain. It is a semantic representation of languages and used as a kind of interlingua among different languages. The most obvious characteristics of the semantic representation are its independence of peculiarities of any language and its underspecification. One parsed source

language is represented with this interchange format, and another target language is generated from it. The whole semantic representation has up to four components: speaker tag, speech act, topic and arguments. The IF has been applied to two languages: Chinese and English. And it is also tested in our translation system.

Section 2 presents the interchange format. Section 3 describes our IF database in the hotel reservation domain. Section 4 gives a brief description of our dialogue translation system. Section 5 discusses the interchange format and makes some comparison.

## 2. THE INTERCHANGE FORMAT

The most obvious characteristics of the semantic representation are its independence of peculiarities of any language and its underspecification. But it must capture the speaker's intent. It is underspecified because it does not define the predicate-argument frame, the voice, the tense, the modality, etc. So we also call the IF as an underspecified semantic representation (USR).

### 2.1 Semantic Representation

The whole semantic representation has up to four components: speaker tag, speech act, topic and arguments as shown in Figure 1. The speaker tag is either "a" for agent or "c" for customer to indicate who is speaking. The speech act indicates the speaker's intent such as giving information, requesting information, expressing thanks and making suggestions. The topic expresses the focus of a sentence. The arguments indicate other information that is needed to express the entire meaning of the source sentence. Both the topic and arguments are made up of attribute-value pairs in functional formalisms. The attribute can be any concept defined in the domain. The value can be an atomic symbol or recursively an attribute-value pair. Both topic and arguments are optional parts in the USR. The symbol "^" in the topic expression indicate that the expression can appear zero to one time, while The symbol "\*" in the argument expression shows that the expression can appear zero to any times. And the attribute-value pairs are order free.

USR ::= speaker: speech act : topic: arguments  
Speaker ::= a|c  
Speech act ::= give-information | request- information | ...  
Topic ::= (concept = attribute)^  
Arguments ::= (concept = attribute)\*

Figure 1: Underspecified Semantic Representation

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

Let us consider several examples. Example (1) is an semantic representation only containing a speaker tag and one speech act without concepts attached to it because the speech act express the enough information of the utterance. In Example (2), the speech act is “request-information” which indicates the customer asks for some information, the topic is “available” and the argument is “time”. In example (3), the semantic representation has no arguments. These three examples present the three types of semantic interchange format in our corpus.

- (1) 您好  
Hello  
a: greet ::
- (2) 我问一下您现在有房间吗  
I wonder if you have rooms left now.  
c: request-information: (available = room): (time = now)|
- (3) 什么时候要哇  
When do you need the room?  
a: request-information :( time = question ):

There are currently 23 different speech acts and 58 concepts in the interchange format. For example, the speech act “give-information” indicates that the speaker wants to inform the hearer of something, while the speech act “request-verification” indicates that the speaker asks the hearer to make some confirmation. Every interchange format must have one speech act. We classified the speech acts into two classes. One is used to control the dialogues. It includes the speech acts that concern with social communications, relate to the dialogue itself (opening or closing the dialogue), or smooth communications. These speech acts include “greet”, “thank”, “introduce” (introducing oneself, including name, title and company), “deliberate”(thinking aloud), and so on. Another is used to conduct the on-going task. These speech acts include “give-information”, “request-verification”, “suggestion”, and so on.

The concepts are organized hierarchically. Some of them can not appear alone. For example, the concept “quantity” must appear together with other concepts such as “room” or “price”. The concepts can be also further classified into two types: general concepts and domain concepts. The general concepts are not relevant to a particular domain and it can be used in any domain. The concepts such as temporal, spatial and location are general concepts. The domain concepts usually only apply to the particular domain. For example, in the hotel reservation domain, the concepts such as room, reservation and hotel are domain concepts. When the IF apply to different domains, we can only change the domain concepts in the IF.

As shown in Figure 1, The topic and the arguments are composed of attribute-value pairs. The topic indicates the focus of the sentence while the arguments include the contents for the sentence to be complete. Values in the pairs are determined by the concepts in the attribute. If the value is recursively an attribute-value pair, the correspondent concept is determined by the concept hierarchy. Only concepts subsumed by the higher concepts can be selected.

We distinguish the topic from the arguments according to the

information structure, which is different from other methods. Any sentence includes two sections: given information and new information. In dialogues, given information is something that has mentioned in the context or something that the hearer has known, while new information is unknown to the hearer. From the context, the given information can be omitted, while the new information cannot. In the new information, there is usually one focus. We represent the focus in the topic and the other information in the arguments. For example, when the customer asks the price of a single room, the price is the focus of the sentence and the type of rooms is only an argument. The interchange format of this sentence is “(c: request-information: (price = question): (room-type = single))”.

In addition, we can also take concepts that need to be emphasized or confirmed as the topic even if the concepts have been touched on earlier in the dialogue. When people asks hearer to make some confirmations, they often use the concept has mentioned before. For example, in the hotel reservation domain, the agent often asks the customer to confirm the number and type of the rooms they have reserved. The typical sentence is “One single room, isn’t it?”. The correspondent IF is “a: request-verification: (room = (quantity=1, room-type =single):”

## 2.2 Undersepcification of the Interchange Format

In order to express the underspecification of the interchange format, first let us consider a complex semantic expression extracted from our corpus. It is shown in Example (4):

- (4) a: give-information: (available = (room = (room-type = double ))) : (price = (quantity = 200&240 , currency = dollar ))|

In Example (4), the speech act is *give-information*, which means that the agent is offering information to the customer. The topic indicates there are double rooms. The arguments list the prices of double rooms, which shows that there are two kinds of double rooms available. So the meaning of this representation is “ We have two kinds of double rooms which cost 200 and 240 dollars respectively”. From the USR, the kinds of rooms are not expressed explicitly in the format. Only from the composite value of the concept “price ” can we judge there are two kinds of rooms because the price is different. This is only one example of underspecification, which needs inferences from the input and the domain knowledge.

From the above example, our semantic representation only expresses the major content critical for understanding. Some linguistic information is not included in the interchange format. The information includes the politeness, formality, tense, modality, anaphora, number, morph etc. The tense can be judged from the speech act or the time word included in the interchange format. The politeness included in the interchange format is not critical for the translation. In different languages, there are different ways to express politeness. When generating sentence in the target language, we can infer the surface form of the sentence from the target language and the speech act. For example, In English people often use “Could you tell me...” when they ask some questions, while in Chinese people

often use “请问...”.

When we use the interchange format, It is more convenient and easy to parse the source language. But it increases the difficulty of generating the target sentences. We must use the domain and common knowledge to infer the missing linguistic information.

### 2.3 Multilinguality

As we discussed above, the interchange format is made up of up to four components: speaker tag, speech act, topic and arguments. The speech act expresses the pragmatic meaning and the interpersonal role of the dialogues. It is independent of a particular language. The concepts in the topic and arguments are semantic ones. General concepts exist in any languages and domain concepts are only relevant to the particular domain. Values of the concepts only list the content and do not express the relationship among them. In other words, the interchange format doesn't concern with the construct of the utterances and the relationship of the constituents in one sentence. It only concerns with the intent and the meaning of the utterance. Obviously, it is independent of particular languages.

In interlingua-based translation systems, the interlingua is one of the key factors which affect the translation results. If the interlingua is too abstract away from the language, the translation result can not express the original meaning of the source sentences. If the interlingua contains too much structure information of the source language, It is difficult to extend to

other languages. In our interchange format, although the IF can not express the whole information of the source information, it can capture the speaker's intent and express the major information critical for understanding. In fact, there exists many conventional ways of expressions for some of the interchange formats. According to the speech act and the content, we can often translate the meaning of the speaker, which does not affect the translation result and, of course, does not affect the understanding between the speakers. So the interchange format we have designed is suitable for domain-oriented speech translation.

### 3. THE IF DATABASE

Our interchange format is first developed in the hotel reservation domain. The whole database includes about 90 dialogues and 3000 sentences [2]. Each conversation turn can be broken into semantic dialogue units. Each semantic dialogue unit corresponds to one semantic representation.

Each unit includes one Chinese sentence, an English translation and a correspondent IF annotated by hand. An example extracted from our corpus is shown in Figure 2:

C: 您要什么样的房间?  
 E: What kind of room would you like?  
 a: request-information: (room-type = question):

Figure 2: A Sample Utterance

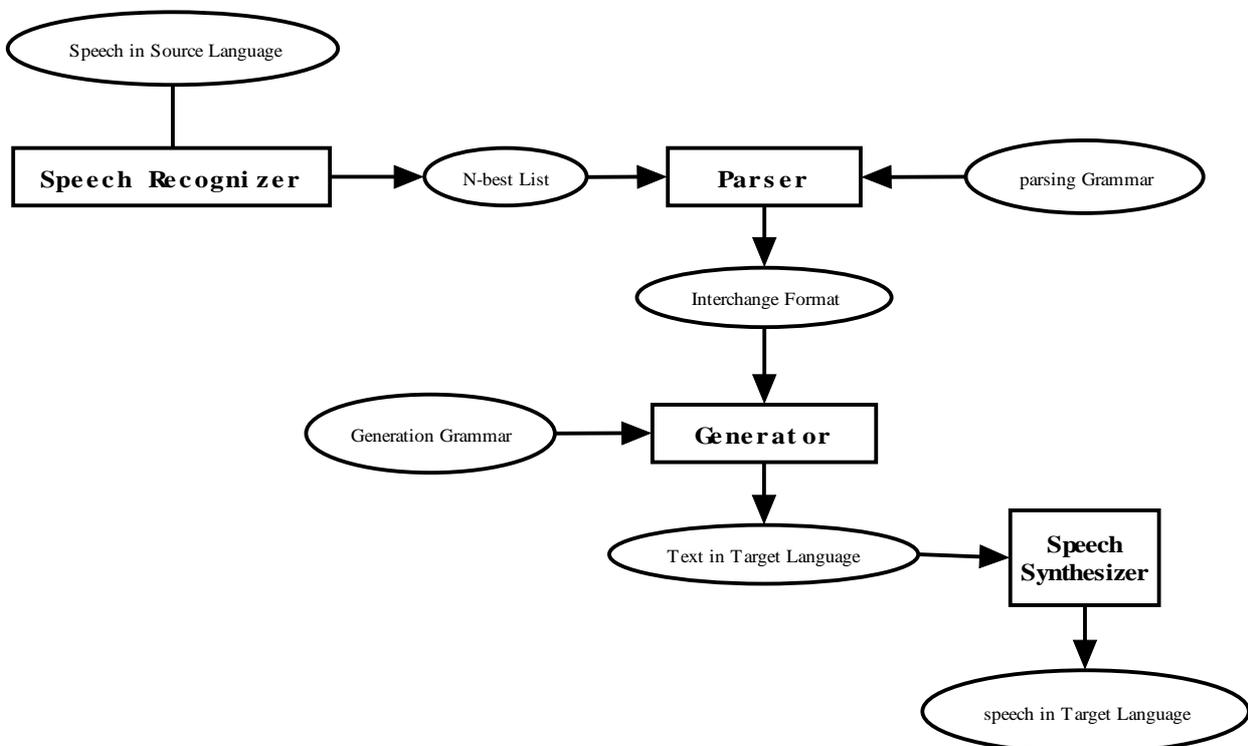


Figure 3: Components of the Translation System

## 4. THE TRANSLATION SYSTEM

The translation system was designed to be a bi-directional system: Chinese to English and English to Chinese. Now we are developing the Chinese to English translator. A component diagram of the translation system can be seen in Figure 3 (A full description of the system is shown in [3]). The main modules of the system are speech recognition, parsing and generation.

The source speech is processed by the speech recognizer into a N-best list of text string hypotheses in the source language. The translator first analyzes the source text into our interlingua semantic representation and then the generator produces a string in the target language. Our parser is a multi-engine one. It is made up of three parts: template matching, rule-based parser and statistical parser. First the input text is matched against the templates. If not matched, the text is passed to the rule-based parser and the statistical parser. Each engine gets a parsing result with a score. Then the two results are compared and get a final result. The IF semantic representation is then passed on to the target language generator. The generator is made of two modules [4]: microplanner and surface realizer. The task of the microplanner is to map the interchange format into the intermediate representation, which is the input of the surface realizer. Then the surface realizer uses the target language grammar to generate a grammatically correct string for the input IF. Currently, there are two generation grammars: Chinese grammar and English Grammar. The speech synthesizer produces the speech in the target language.

Now the interchange format has been tested off-line using our generator. Tests show that 90% of the generated sentences are rated as grammatically and semantically correct. The other 10% are rated as wrong because the mood of the sentences is not correct [4]. This is mainly caused by the lack of the dialogue context and the mood not expressed explicitly in the format.

## 5. DISCUSSION

The interchange format is first developed under the guide of our hotel reservation domain. Then it is extended to the travel information domain. Result of annotation by hand shows that the IF can cover almost all the corpus.

The C-STAR consortium has also developed an interchange format for speech translation [5][6]. The basic frame is similar to us. But there are more constraints on the topic and arguments. The topic can be made up of several concepts. And the order of these concepts is defined. Any concept list violating this order is considered as illegal. And the arguments are determined by the speech act and the topic in the IF. Our topic is a concept-attribute pair and it is the focus of the sentence. All other information is left to the argument part and they are order free. This made the IF simple and useful for parsing and generation.

The IF is constructed according to our corpus. The speech acts express the speaker's intents. It also needs to be extended when it is applied to other domains because different speech

acts are needed in different domains. Although the IF conducts well in our translation system, it lacks some information mentioned in section 2.2 and sometimes causes translation errors [4]. We plan to study the problems caused by it and see how these problems can be handled in the future.

## 6. REFERENCES

- [1] David R. Traum, *Speech Acts for Dialogue Agents*. In Michael Wooldridge and Anand Rao, editors, "Foundations And Theories Of Rational Agents", Kluwer Academic Publishers, pages 169--201, 1999.
- [2] Chengqing Zong, Hua Wu, Taiyi Huang, Bo Xu. *Analysis on Characteristics of Chinese Spoken Language*. In the Fifth Natural Language Processing Pacific Rim Symposium, pp358-362, 1999
- [3] Chengqing Zong, Taiyi Huang, Bo Xu. *Design and Implementation of a Chinese-to-English Spoken Language Translation System*. ISCSLP, Beijing, China, 2000.
- [4] Hua Wu, Taiyi Huang, Chengqing Zong, Bo Xu. *Chinese Generation in a Spoken Dialogue Translation System*. COLING2000, Saarbrucken, Germany, 2000
- [5] Lori Levin, Donna Gates, Alon Lovie, Alex waibel. *An Interlingua based on Domain Actions for Machine Translation of Task-Oriented Dialogues*. ISCLP 98, Sydney, Australia.
- [6] L.M. Tovina. *A user's point of view on the interlingua IF*. ITC-IRST Technical Report 9809-01, 1998