In Proceedings of ICSLP-98

An Interlingua Based on Domain Actions for Machine Translation of Task-Oriented Dialogues

Lori Levin, Donna Gates, Alon Lavie, Alex Waibel

Language Technologies Institute
Carnegie Mellon University
email : lsl@cs.cmu.edu

ABSTRACT

This paper describes an interlingua for spoken language translation that is based on *domain actions* in the travel planning domain. Domain actions are composed of speech acts (e.g., request-information), attributes (e.g., size, price), and objects (e.g., hotel, flight) and can take arguments. Development of the interlingua is guided by a database containing travel dialogues in English, Korean, Japanese, and Italian. There are currently 423 domain actions that cover hotel reservation and transportation. The interlingua will soon be extended to cover tours, tourist attractions, and events. The interlingua is used by the C-STAR speech translation consortium for translating travel planning dialogues in six languages: English, Japanese, German, Korean, Italian, and French. The paper also addresses the role of the interlingua in Carnegie Mellon's JANUS translation system.

1. INTRODUCTION

Task oriented domains such as travel planning include a large percentage of formulaic utterances that cannot be translated literally. For example, *Could you tell me the departure time?* doesn't make sense when translated literally into Japanese as *Shuppatsu jikan wo (depart time) watshi ni (to me) iemasu (can tell) ka (?).* Instead it would be more appropriate to say something like *Shuppatsu jikan wo (depart time) oshiete (teaching) itadakemasen (couldn't receive) ka (?).* Translating such utterances requires knowing the speaker's intention or speech act, such as requesting information in this case. Taking the notion of speech acts one step further, we can identify *domain actions* such as requesting information about a flight time or giving information about the price of a room.

This paper describes an interlingua for machine translation of spoken travel planning dialogues that is based on such domain actions. The interlingua, known as IF (Interchange Format), is used by C-STAR, a multi-national consortium of research groups collaborating on speech-to-speech translation. The C-STAR languages are Japanese, English, German, Korean, Italian, and French.[1] C-STAR adopted an interlingua in order to facilitate translation between as many language pairs as possible with minimal effort. Sites that wish to use IF supply an analyzer that produces IF from sentences in the home language and a generator that takes IF as input and produces sentences in the home language. Using the analyzer from one language and the generator from another results in translation from the first to the second language.

[1]We would like to thank our sister lab at the University of Karlsruhe, Germany and our other partners in the C-STAR Consortium who have collaborated with us on the design of the interlingua: ATR, Japan; ETRI, Korea; IRST, Italy; CLIPS, France; and Siemens, Germany.

2. THE INTERCHANGE FORMAT

The most important factor in the design of the IF is that it must abstract away from peculiarities of any particular language in order to allow for translations that are non-literal, but capture the speaker's intent. As mentioned above, in the travel planning domain non-literal translations may be required because of many fixed expressions that are used for activities such as requesting information, making payments, etc. An additional factor that constrains interlingua design is that it is used at multiple research sites. It was therefore necessary to design a simple interlingua that could be used reliably by many MT developers with greatly varying translation systems. Simplicity is possible largely because we are working on travel planning, a task-oriented domain with clearly identifiable domain actions (DAs). These domain actions are the basis of the IF. The remainder of this section describes the structure of DAs.

Each DA has up to four components: the *speech act*, the *concepts*, the *arguments*, and a *speaker tag*. Plus signs separate speech acts from concepts and concepts from each other. In general, each DA has a speaker tag and at least one speech act optionally followed by a string of concepts and optionally, a string of arguments. DAs can be roughly characterized as shown in (1). However, there are constraints on the order of concepts so that not all combinations are possible.

(1) $speaker : speech\ act + concept^* \ argument^*$

In example (2) the speech act is `give-information`, the concepts are `availability` and `room`, and the arguments are `time` and `room-type`. The possible arguments of a DA are determined by inheritance through a hierarchy of speech acts and concepts. In this case `time` is an argument of `availability` and `room-type` is an argument of room. Example (3) shows a DA which consists of a speech act with no concepts attached to it. The argument `time` is inherited from the speech act `closing`. Finally,

example (4) demonstrates a case of DA which contains neither concepts nor arguments. The following paragraphs describe the four components of DAs, speaker tags, speech acts, concepts, and arguments.

(2) On the twelfth we have a single and a double available.
`a:give-information+availability+room (room-type=(single & double),time=(md12))`

(3) And we'll see you on February twelfth.
`a:closing (time=(february, md12))`

(4) Thank you very much
`c:thank`

Speaker Tag: The speaker tag is either `a:` for agent or `c:` for customer to indicate who is speaking. The speaker tag is sometimes the only difference between the IFs of two different sentences. For example, *Do you take credit cards?* (uttered by the customer) and *Will you be paying with a credit card* (uttered by the agent) are both requests for information about credit cards as a form of payment.

Speech Acts: There are currently 38 speech acts defined in the IF. Some speech acts are very general. For example, `give-information` is used in many DAs where the speaker's intent is to inform the listener of something, such as `give-information+temporal+departure+flight`, `give-information+expiration-date`, etc. Others are more specific, such as `delay-action`, which is used specifically for utterances like *I'll get back to you on that*. Normally each DA has one speech act. However, there are three special speech acts that combine with other speech acts. These are `verify`, `request-verification`, and `negate`. For example the sentence *So you're not leaving on Friday, right?* has the speech act `request-verification-negate-give-information`.

Concepts: There are currently 68 concepts defined in the IF. Each DA can have zero or more concepts following the speech act, although not all possible strings of concepts are allowed. Concepts fall into several classes that roughly constrain how they combine with each other. Some classes of concepts are actions (change, reservation, confirmation, cancellation, etc.), attributes (availability, size, temporal, price, location, features, etc.), and entities (room, hotel, flight, numeral, expiration date, etc.). The usual order of concepts in a DA is `action+attribute+entity` as in `request-action+reservation+temporal+room` for *I'd like to make a reservation for a room on the fifth*. In this case, the speech act is `request-action` and the concepts are `reservation`, `temporal`, and `room`.

The concept components of a DA capture the focus of a sentence. For example, the sentence *The week of the twelfth we have both singles and doubles available* mentions a date, a room type, and the notion of availability. However, since the focus of the sentence is availability, the dialogue act is `a:give-information+availability+room` and the time and room type are expressed as arguments of this dialogue act.

Arguments: Arguments add specific information to the DA, such as times, prices, and specific features of entities. An argument consists of an argument name and a value separated by an equal sign, for example `room-type=double`. In addition to atomic values, there are various types of complex values as shown in examples (5)-(13). Multiple values and coordination can combine with price, time, interval, frequency, and duration for arguments like *on July 5 and July 6 at 4:00*.

(5) **multiple values:**
`room-type=(double,non-smoking)`
a non-smoking double

(6) **coordination:**
`room-type=(single & double)`
a single and a double

(7) **quantity:**
`room-type=(double, quantity=2)`
two doubles

(8) **price:**
`price=(currency=dollar, quantity=50, per-unit=night)`
fifty dollars per night

(9) **time:**
`time=(md5, tuesday, july, 1998, 16:00, afternoon)`
Tuesday July 5, 1998 at 4:00 in the afternoon

(10) **time interval:**
`time=(start-time=(md5, july), end-time=md10)`
from July 5 to 10

(11) **duration:**
`duration=(time-unit=day, quantity=9)`
for nine days

(12) **frequency:**
`frequency=(time-unit=hour, quantity=2)`
every two hours
`frequency=(per-unit=hour, quantity=2)`
two times per hour

(13) **lists of characters:**
`spelling=[g, a, t, e, s]`
g a t e s

The possible arguments of a DA are determined by the speech acts and concepts it contains. For example, `give-information+temporal+flight` can take the arguments associated with the concepts `temporal` (time, duration, frequency) and `flight` (flight-type, carrier-name, flight-number, destination, origin). There are currently 86 argument names defined in the IF.

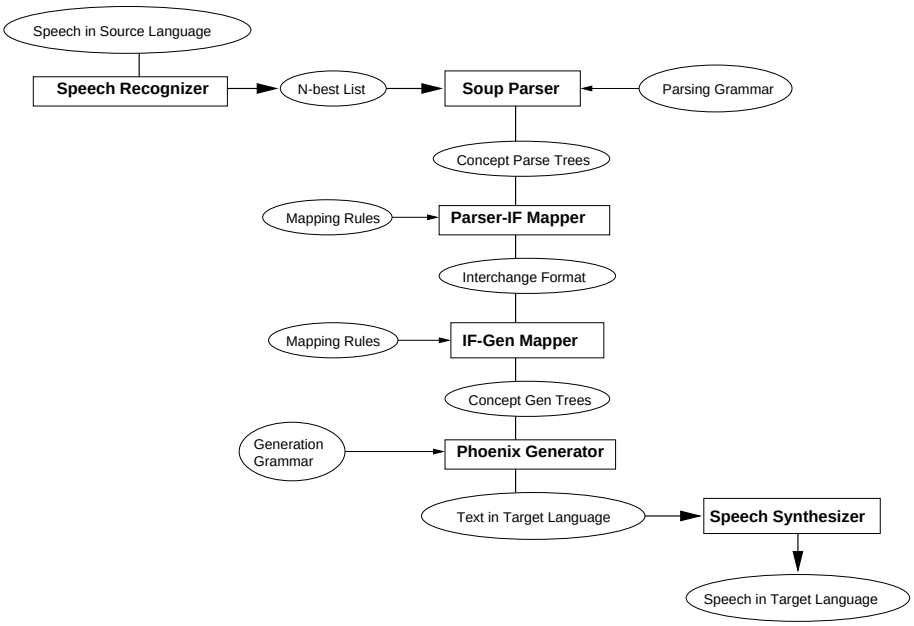

Figure 2: Components of the Translation System

```
61.2.3  olang I  lang I  Prv IRST
‘‘telefono per prenotare delle stanze per me
  e quattro colleghi’’
61.2.3  olang I  lang E  Prv IRST
‘‘i’m calling to book some rooms for me and
  four colleagues’’
61.2.3  IF  Prv IRST
c:request-action+reservation+features+room
(for-whom=(I & (associate, quantity=4)))
61.2.3  comments: dial-005-spkB-roca0-02-3
```

Figure 1: A Sample Entry from the C-STAR IF Database

	Dialogues	SDUs
English	32	2400
Korean	70	1140
Italian	5	233
Japanese	124	5887

Table 1: Size of the Dialogue Act Database

3. THE IF DATABASE

Development of the IF is guided by an IF database containing travel dialogues in English, Korean, Japanese, and Italian. The size of the database is summarized on Table 3. Conversational turns in the database are broken down into *semantic dialogue units* (SDUs). SDUs are sentences that correspond roughly to a DA.

Each entry in the database represents one SDU. An entry contains an SDU in one of the C-STAR languages, an English translation, an IF, and possibly some comments. The entry also specifies the dialogue number and utterance number that the SDU came from, the original language, and the provider. An example of a database entry is shown in Figure 2 This is SDU 3 from utterance 2 of dialogue 61 provided by IRST in Italian.

Although there are 268,816 allowable combinations of speech acts, concepts, and speaker tags, only 423 actually occur in the IF database.

4. THE JANUS-III TRANSLATION SYSTEM

The JANUS-III MT system was designed to accommodate multi-party, multi-lingual conversations between travellers and travel agents. A component diagram of the JANUS speech translation system for the travel domain can be seen in Figure 2. The main system modules are speech recognition, analysis, and generation. The analyzer and generator are language-independent in that they consist of a general processor that can be loaded with language specific knowledge sources. Our travel domain system currently includes analysis grammars for English and German and generation grammars for English, German, and Japanese.

The interface between the speech recognizer and the translation system is via an N-best list of text string hypotheses in the source language. Translation is then performed by analyzing the text string in the source language into our interlingua representation, and then generating a string in the target language. First, the input string is analyzed by SOUP [3], a robust parser designed for spoken language. SOUP works with semantic grammars in which the non-terminal nodes represent concepts and not syntactic categories. The Parser-to-IF mapper then converts this representation into the

canonical *Interchange Format* representation described earlier in the paper. The mapper performs a simple format conversion, and does not contribute any significant information beyond that derived by the parser. The IF interlingua representation is then passed on to generation. The generation process first uses a generation mapper, which converts the IF into a tree semantic representation. The tree is then passed on to the PHOENIX [4] generator, which can generate output text for several different target languages (currently English, German and Japanese) using target language generation grammars. Note that this framework supports generation back into the source language (in our case, English), which results in a paraphrase of the input. This provides the user with a mechanism for verifying analysis correctness, even when he/she is not fluent in the target language. The IF can also be exported to the generation systems of other C-STAR partners for translation into languages not supported at CMU (French, Italian, and Korean).

In addition to the system described here, we are also experimenting with alternative approaches that will complement and increase the robustness of the system. These include a statistical method for assigning DAs to sentences ([2]) and a direct glossary-based translation approach ([1]).

5. COVERAGE AND LIMITATIONS

The C-STAR consortium is focusing on four sub-domains of travel planning – hotel reservation, transportation (plane and train only), sight seeing, and events (e.g., festivals and sporting events). The IF is fairly well developed for hotel reservation. In an informal test on two previously unseen dialogues containing a total of 143 SDUs, our project linguists determined that our current set of DAs and arguments covers 92% of the SDUs. The IF is also fairly well, but slightly less developed for transportation. We have just begun development on sightseeing and events. The IF also covers cross-domain phenomena such as greetings, closings, and other phrases that are involved in conversation management.

Many linguistic phenomena are not covered by the IF. Phenomena that are not covered include comparatives, relative clauses, extensive noun modification, modality (possibility, necessity, etc.), politeness, formality, certainty, tense, aspect, connectives (e.g., *because*) between sentences, anaphora, and number. Some of this information will need to be added in the future. For example, comparatives and relative clauses do occur in our data. Other features like tense are often (but not always) predictable from the DA. And others like modality (e.g., the *could* in *Could you tell me . . .*) are generally part of the formulaic, conventional ways of expressing the DAs in specific languages, but their form is not relevant for translation. For example, in parsing English *could* helps to identify a request, but it should not appear in the IF because it does not translate literally into other languages.

We have also come across some cultural problems with translation. The opening phrase *How can I help you?* is not used by Japanese service providers, and sounds strange when translated into Japanese, but if it is replaced by another expression or left out then the flow of conversation would be interrupted on the English side. Similarly, the closing *Thank you for using World Wide Travel* does not sound appropriate when translated into German. These problems are not specific to the IF-based approach. We are planning to study protocols for multi-cultural dialogues in order to see how culture-specific conventions should best be handled.

6. REFERENCES

1. Frederking, Robert, S. Nirenburg, D. Farwell, S. Helmreich, E. Hovy, K. Knight, S. Beale, C. Domashnev, D. Attardo, D. Grannes, and R. Brown: 1994, 'Integrating Translations from Multiple Sources within the Pangloss Mark III Machine Translation', *Proceedings of the First Conference of the Association for Machine Translation in the Americas* (AMTA-94). Columbia, Maryland.
2. Fukada, Toshiaki, Detlef Koll and Alex Waibel: 1998, "Probabilistic Dialogue Act Extraction for Concept Based Multilingual Translation Systems," ICSLP 98, Sydney, Australia.
3. Gavaldà, Marsal: 1998, The SOUP Home Page. `http://www.is.cs.cmu.edu/ISL.speech.parsing.soup.html`
4. Ward, Wayne: 1990, 'The CMU Air Travel Information Service: Understanding spontaneous speech', In *Proceedings of the DARPA Speech and Language Workshop*.