MEANING EXTRACTION BASED ON FRAME REPRESENTATION FOR JAPANESE SPOKEN DIALOGUE

Akira KUREMATSU, Takeaki Nakazaki
Course of Electronics Engineering, Univ. of Electro-Communications
Chofu-shi, Tokyo 182-8585, JAPAN
e-mail: kure@apple.ee.uec.ac.jp

ABSTRACT
This paper describes the issue of meaning extraction based on frame representation for spoken dialogue. The framework of semantic lexicon and the structure of semantic caseframe which focuses on the keywords in phrase is described. The results of the preliminary experiments of meaning extraction using spontaneous speech in the scheduling task and the document retrieval task are shown.

1 INTRODUCTION
In the practical application of speech recognition, the task domain is restricted. It means that the semantic meaning of uttered words and the syntactical and semantic knowledge are constrained in the specific domain. Recognition of spoken dialogue introduces serious problems, as spontaneous speech contains grammatically ill-formed expressions imperfect or incomplete utterances. The promising way of the understanding of spontaneous speech is the extraction of the meaning of the utterance, not the correct recognition of each uttered word. Semantic information is effective for spoken language understanding in the limited task domain. It is important to determine the meanings of word sequences in the understanding of spoken dialogue. Spontaneous speech utterance is converted into ill-formed word sequence. Case systems as a means of meaning representation for natural language has been discussed extensively [1]. Case has meant the classification of noun forms according to their inflection. Especially, deep structures for natural language utterances and storage structures for artificial language programs have been discussed. However, meaning representation derived from shallow processing of obtained word sequence can be expected as the practical method for the spoken dialogue understanding [2]. The technique underlying the application of linguistic knowledge and semantic expectations is semantic caseframe instantiation [3]. Since the more significant words in Japanese dialogue are verbal words and nouns in relation to them, the keyword extraction as well as the syntactic and semantic information will be the appropriate way of spontaneous speech and meaning processing.

2 MEANING EXTRACTION

2.1 Semantic Caseframe
The caseframe method enables the extraction of the meaning by anchoring its interpretation on the parts of the input that are most significant, and referring to the related less significant objects. The central notion behind the caseframe for Japanese utterance is that of a head concept extracted in the verb or predicate part modified by a set of related cases.

The semantic caseframe can be applied to the mapping of each action in the domain onto a caseframe. The domain specific caseframe has cases for each of the domain objects that are involved in the associated actions. The caseframe is identified as a verb or clausal caseframe corresponding to the verb (Headforms). In addition to the caseframe associated with actions, the caseframe to represent objects where cases correspond to modifiers are also used. This has essentially the same form as the verb caseframes, except that its headforms correspond to the head nouns of a noun phrase.

2.1 Semantic Lexicon
To apply a caseframe for meaning representation, a semantic lexicon is to be developed based on the surface semantics. Surface semantics is the knowledge about the usage of words in contexts. The semantic dictionary has been formed to contain the word category and the possible case attributes. The attributes are described using the Japanese thesaurus classification [4].

Each word contains the typical attribute values. The verbal semantic attributes have been created in consideration of the semantic relationship between verbs. As for the semantic representation of nouns, conceptual meaning has been classified into about 3000 attribute categories of primitive meaning. The semantic categories
are classified into hierarchical tree structure in terms of the is-a relationships and the has-a relationships. The depth of categories amounts to 12 at the deepest level. Example of semantic structure of verbal patterns is shown in Figure 1:

**Japanese Verb: aku ``be free``**

N1 (SUBJECTS or TIME)

``Doyoubi ga aite i masu." (Saturday is free.)

Verbal semantic attribute: N1's attribute transfers or N1's attribute is moved.

**Japanese Verb: iku ``go``**

N1 (SUBJECTS or PEOPLE)

N2 (PLACE)

``Watashi ga sochira no heya ni iki masu." (I shall go to your office.)

Verbal semantic attribute: N1's physical action transfers to N2.

**Figure 1:** Example of semantic structure of verbal patterns

### 2.2 Semantic Frame

In extracting meaning of the sentence, the attributes of semantic frame are represented based on the verbal headform. The structure of the frame is composed of the attributes expressed in the semantic lexicon of the verbal word. The term attributes of semantic lexicon contains the orthographic representation, termination form of verbal words, part-of-speech, the semantic category, cases and particles to be related. The matching of the attribute items in the dictionary and the instances in the input word candidates were processed to extract the meaning of the sentence.

The matching of the attribute items in the dictionary and the instances in the input word candidates are processed to extract the meaning of the sentence.

In the semantic lexicon of verbal words, multiple part-of-speech classes of meaning are expressed, if there exist. Moreover, what kind of case particle attached to a noun word before or after the verbal word is also associated. The semantic category is referred based on the Japanese semantic dictionary. Therefore, it is possible to extract meaning from the feature of noun even when the case particle is not correctly recognized.

For example, the most important part of the utterance in the scheduling task is the appointment of the schedule. The examples of the significant words were extract as "have time", "be occupied". The relationship of the objects like "of (/no/)", "from (/kara/)", "to (made/)", "and (to/)", "or (/matawa/)" are also extracted by the analysis of the particle. The example of frame of verbal word of "be free, be occupied" in the scheduling task is shown in Figure 2.

![Figure 2: Example of frame of verbal word](image)

### 2.3 Meaning Relationship

The semantic frame is instantiated from the analysis of the case of Japanese particle of the dialogue sentence. The meaning relationship of particles (in Japanese) is shown in Table 1.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Example</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Of-rel (/no/)</td>
<td>Three o'clock of tomorrow</td>
<td>(Three o'clock, (of: tomorrow))</td>
</tr>
<tr>
<td>From-rel (/kara/)</td>
<td>From three o'clock</td>
<td>(from: three o'clock)</td>
</tr>
<tr>
<td>To-rel (/made/)</td>
<td>To three o'clock</td>
<td>(to: three o'clock)</td>
</tr>
<tr>
<td>Quant-rel (/wa/)</td>
<td>Monday is three o'clock</td>
<td>(three o'clock), (quant: Monday))</td>
</tr>
<tr>
<td>And-rel (/to/)</td>
<td>Monday and Tuesday</td>
<td>(Tuesday, (and: Monday))</td>
</tr>
<tr>
<td>Or-rel (/matawa/)</td>
<td>Monday or Tuesday</td>
<td>(Tuesday, (or: Monday))</td>
</tr>
</tbody>
</table>

The example of "and-relationship" based on the Table 1 is shown in the following. From the word sequence of "Monday and Tuesday", the particle "and" is recognized and the classes of words before and after the particle are referred in the semantic dictionary. The following expression is obtained. N indicates the noun class.

(N: Monday, (and: N: Tuesday))

### 3 EXPERIMENTS

#### 3.1 Scheduling Task

**1) Corpus Description**

In the scheduling task, two speakers talked each other looking at separate calendars. The speaker used one of 25 two or four week calendars filled in his/her appointment schedules. The utterance in this scenario followed the
calendar constraints to use the words indicated in the appointment. The speaking style was free and the utterances were spontaneous. The average period and number of turns were 2.48 minutes and 14.14, respectively. The speakers sat in the same room but do not face each other in order to avoid non-verbal communication. The speaking style was the controlled one where the speaking turn was controlled and the overlapping of utterances was reduced.

The collected dialogues were transcribed in Kanji-Kana sentence and Roman script sentence. Both speech and non-speech events were transcribed. Colloquial pronunciation was transcribed as it was heard and the normal expression was shown in parentheses. The speech was transcribed including words, human noises, non-human noises, silence, false starts, and transcriber comments.

In Roman transcription, phrases were separated by space. Word segmentation was undertook based on the morphological program of CHASEN [6]. The segmentation was obtained based on the word occurrence and DAIJIRIN dictionary [7]. Part-of-speech was tagged by use of morphological analyzer. Sample of the transcription is shown in Figure 3.

Figure 3: Sample of transliteration of dialogue utterances

A: /eetoo/ <P> raishuu <P> de are ba <P> /eel/ <P> kayoo(&kayou) to kiNyoobi(&kiNyoubi) <%wa> ichinichi (&ichinichi) ai ta/= ai te ori masu keredomo . <P> seNsee(&seNsei) <P> go tsugoo(&tsugou) ikaga desu ka .

(2) Characteristics of Corpus

The total number of dialogues of the scheduling task was 800 and the number of sentence was 21,627. As the results of statistical analysis of the scheduling task, there were 3,359 different words including various noises. The number of different words excluding noises and mispronunciation was 2,742. Testset perplexity was 42.0 using bigram as a language model. Some characteristics of spoken dialogues were that the ending of the sentence was not completed from a grammatical point of view. That is because the ending was noun, or in the middle of sentence. Long sentence in the controlled dialogue, where conjunctive particle of such as /keredomo/(though, although), /ga/(but, however), /de/(and, then) combine short sentences. Frequent colloquial lazy pronunciations appeared.

(3) Speech Recognition System

The outline of speech recognition system which we used for this experiment is as follows. As an acoustic model, HMM phoneme model of continuous mixture was used. As a language model bigram was used. The correct word recognition rate of 79.05% was obtained in the open test of 691 utterances. Noise recognition was performed by use of HMM noise model. The value of 86.9% of noises including pause were discriminated correctly from speech, where 23.96% of whole words were either noises or pause.

(4) Experiments anf Results

The experiment of meaning extraction was undertaken for the data in the scheduling task. The output of speech recognition was analyzed by use of the Japanese morphological analyzer and the predicate part was extracted. By referring the semantic dictionary, the instance was transformed to object of caseframe and the content of caseframe was instantiated. The semantic attributes are the extraction of information on time, date and time from the categorical meaning of recognized words and their part-of-speech. The important part of the utterance is the appointment of the schedule and the examples of the significant words were extract as "have time", or "be occupied". The relationship of the objects like "of", "from", "to", "and", "or" were also extracted by the analysis of the Japanese particle. The predicate verbs were extracted with frequent occurrence. The semantic analyzer was developed to extract the semantic caseframe. The output of speech recognition was analyzed by use of the Japanese morphological analyzer and the predicate part was extracted. By referring the semantic dictionary, the instances were transformed to objects of caseframe and the contents of caseframe was instantiated. The performance of the meaning extraction was evaluated by the scoring of the correct extraction of predicate and objects of the caseframe for each sentence. The 66% of the meaning extraction score was obtained for the learning data and the 86% was obtained for the test set data.

Parsing program to extract meaning based on the frame representation of verbal word was produced and the experiment of recognition and meaning extraction was performed by use of the spontaneous dialog corpus. The 15 sentences were extracted for the development of semantic dictionary. The evaluation test was performed for the open data of 5 phrases of the different dialog of scheduling task. The predicate were five frequent verbs. The five most frequently occurred verbs were chosen in the experiment. In the open test, the vocabularies
The performance of the meaning extraction was evaluated by the scoring of the correct extraction of predicate and objects of the caseframe for each sentence. The 66% of the meaning extraction score was obtained for the learning data and the 86% was obtained for the test set data.

3.2 Document Retrieval Task

The second experiment was undertaken for spoken dialogue in the task of document retrieval. In this case, the technical paper is searched from the database by use of the spoken input for retrieval. Data was collected by Wizard-of-Oz method. We used 36 dialogues of the typical inquiry expression for this experiment. The frame components included the action referring to verb and names of author(s), year, name of the institute, content of the document, and category field. As for the speech recognition system, the speech recognition engine based on HMM net was used [8]. The results of speech recognition was applied to the process of meaning extraction. The semantic analyzer was modified to meet the conditions of the frame components of this task. When the input was speech, 88% of the meaning extraction score was obtained for the learning data and 68% was for the test set data, respectively. In the case of the text input of utterance, the score was 92%.

4 DISCUSSIONS

(1) From the results of experiments, the proposed procedure of meaning extraction seems to be a promising method which will be applied to Japanese spoken dialogues of various task domains. It was shown that the meaning extraction was possible without being influenced by the deviations of speech in the spontaneous utterances. The availability to extract the meaning can be expected by use of the parsing program. In the experiment of the scheduling task, the high extraction rate was obtained, since the kinds of verbs were limited into a few numbers in the limited speech data.

(2) Since there are inflations of verb and the stem part is usually short in Japanese, the special consideration is required when the numbers of verbs are increased.

(3) In these experiments, the segmentation of phrases were conducted by manual operation. It will be necessary to segment phrases properly by extracting the keywords to segment the phrase automatically. When the speech of the keywords is not possibly recognized or when no such keywords appear, it will be effective to use the method to segment input speech into phrases by use of prosodic characteristics.

(4) The recognition of date information in the scheduling dialogue is very important. The conjugate of number will be appropriately treated as one compound word in stead of separated words in order that the numbers are correctly recognized.

(5) The experiment of meaning extraction by use of a large volume of data will be required to evaluate the effectiveness of this method.

5 ACKNOWLEDGEMENT

The author gratefully acknowledge the support of the Japan Society for the Promotion of Science. We wish to express my thanks to members of intelligent electronics laboratories at UEC.

6 REFERENCES