ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

An Example-based Approach to Semantic Information Extraction from Japanese Spontaneous Speech

Otoya SHIROTSUKA and Ken'ya MURAKAMI
Laboratory for Information Technology, NTT DATA Communications Systems Corporation
Kawasaki, Kanagawa 210, Japan

Abstract

Dealing with the linguistic phenomena of spontaneous speech by the existing rule-based approach requires the preparation of complex analysis rules, which takes a great deal of effort. This paper describes a new method of extracting semantic information extraction from Japanese spontaneous speech by an example-based approach (EBA). Compared to the rule-based approach, EBA is robust and requires little effort for knowledge acquisition and its formation. In experimental evaluations of a semantic information extractor based on EBA, transcriptions of one hundred spontaneous dialogues are used as an example corpus and a testing corpus. The best performances of the extractor are 81.6% for precision rate and 62.2% for coverage rate in semantic feature extraction. The results suggest that our method is robust against unknown words and ill-formed sentences, and the extractor proved that EBA can be used as an effective tool foro extracting semantic information from spontaneous speech.

1 Introduction

The semantic analysis of spontaneous speech poses many problems. The existence of ungrammatical sentences, unlimited range of vocabulary, strong dependencies on dialogue context, etc, all pose difficult technical hurdles. To extract semantic information from Japanese spontaneous speech, we must deal with the linguistic phenomena of spontaneous speech [1].

A rule-based analysis for spontaneous speech requires complex grammatical and semantic rules and robustness to deal with such linguistic phenomena. This implies that a great deal of effort is required in order to construct these analysis rules.

In the field of machine translation (MT), the difficulty of building large-scale rule bases is already recognized. One approach to this problem is the example-based approach. Example-based machine translation (EBMT) [2] [3] utilizes a database of examples (pairs of source text and its translation) as the knowledge. In this approach, translation is performed by retrieving similar examples selected from the example database and adapting them in order to translate the input sentences. The performance can be improved simply by adding more appropriate examples to the database. Since it requires no rules and the use of an example is relatively localized, EBMT can be upgraded more easily than rule-base machine translation (RBMT).

Focusing on the features of EBA in MT, we applied EBA to the extraction of semantic information from Japanese spontaneous speech and investigated its ability.

2 Spoken Dialogues

2.1 Data

One hundred spontaneous Japanese dialogues between a customer and a clerk in charge of scheduling conference rooms were collected and transcribed. The fifty people who played the part of customer were informed about their situation context in advance. The transcriptions of customers' speech were used as example and testing data. Speech data was collected using four microphones: two head-mounted microphones and two microphones with stand (unidirectional for the customer and omnidirectional for both).

In the spontaneous spoken dialogues, it is difficult to define an "utterance" since there are many interruptions by the other speaker, so two utterances often overlap. In this paper, an utterance is transcribed as a segment of speech that runs from the word after the final word of the previous utterance to the word before the partner's utterance starts.

For example, during the production of the customer's speech "getuyou ni yoyaku site kudasai", meaning: [please reserve on monday], if the clerk says "hai"[yes] just after the customer's word "ni"[at] as a word of agreement, the transcription of customer's speech should be two separate sentences: "getuyou ni"[on monday] and "yoyaku site kudasai"[please reserve].

2.2 Linguistic Features of Spoken Dialogues

Below we list some typical types of linguistic phenomena in spontaneous dialogue with examples observed in transcribed spoken Japanese dialogues.

False start:

- yoya yoyaku site kudasai [reservation please]

'yoya' is an incomplete fragment of the next word 'yoyaku' [reservation].

Repairs:

- one yoyaku site kudasaidai [reservation please]
  The word 'one' is a fragment of the word 'onegaisimasu' [please].
- yotee yotee sitemasu [(I) have a plan]
  The word 'yotee' [plan] was repeated.
- dai iti haado ua uea tantou no
  [of 1st hardware (group)]

The phrase 'haado ua uea' is a poor pronunctiation of 'haadouea' [hardware].

Syntactic error with no repair:

- yoyaku ga onegaisimasu [please reserve]
- kaigisitu wo yoyaku oboeteinai no desuga
  [(I) don't remember the reservation of the conference room]

These are syntactically incorrect utterances. In the first example, the case particle 'ga' should be 'wo'. In the second example, the case particle 'wo' should be 'no'.

Incomplete utterance by interruption:

- kyuu kai no kaigisitu wo onegai...
  [(please reserve) a room of 9th floor]
- ...no tutumida desu [of Tutumida]

These are utterances which have incomplete or missing parts as a result of an interruption of the clerk's utterance.

Filler phrase with no information:

- sou desu ka, eeto desu ne, ano
  [oh well], [let me see], [uh...]

These are frequently inserted phrases containing no significant information.

Fragmental Speech:

- getuyou, dai sann de iti zi kara, ookii hou
  [monday], [with the third], [larger one]

These are minimal forms of utterances observed in the dialogues generally.

2.3 Semantic Information to Extract

The final goal of this study is to develop a new method of extracting semantic information that is robust against linguistic phenomena of spoken dialogues, for a human-computer interaction system that enables cooperative work through dialogues. As a first step of the study, the semantic information to be extractd is limited to the sphere of information that is necessary to complete the domain and can be obtained from the local context within the sentence itself. This semantic information is called a list of semantic features. The following is an example of a typical sentences in the domain corpus and a corresponding semantic feature list.

- eeto desu ne zikan wa desu ne ni zi kara san zi kan yoyaku site kudasai

  'uh well, time is from 2 to 4 o'clock, make reservation please'

```
y: time
   time: 1400
   senseform: time_from
   time: 3
   senseform: time_interval_of
   x: reserving(action)
```

In the domain of conference room scheduling, 21 semantic feature categories and 117 semantic features are defined.

3 EBA Semantic Feature Extraction Algorithm

Our EBA semantic feature extraction processes input sentences in two stages: sentence-based processing and phrase-based processing.

Sentence-based processing:
The most similar example sentence is located within the example corpus. The similarity L of sentences is calculated as follows:

$$L = \frac{\min_{i=1}^{n} D(I, C_i)}{len(C_i)}$$

where
I : input sentence
C_i : example sentence i
$D(I, C_i)$: distance between I and C_i
n : total number of example sentences
$len(C_i)$: total number of words in C_i

Phrase-based processing:
Phrase alignment information between phrases of an example sentence and semantic features is prepared in advance. Phrases of the input sentence are

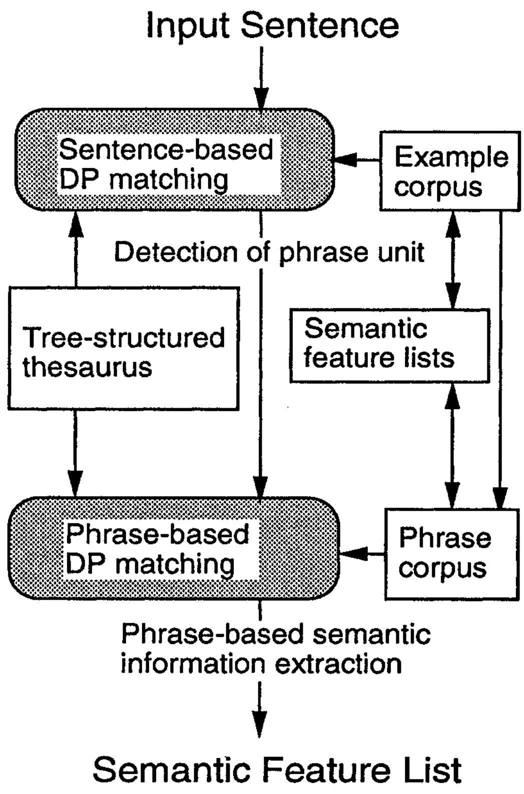

Figure 1. **Flow of the Extraction Process**

detected by sentence alignment information between the input and the selected sentence, and the phrase alignment information. Then a semantic feature list of input sentence is constructed by collecting semantic features corresponding to the example phrases that are most similar to the detected phrases of the input sentence.

Figure 1 shows the flow of the the extraction process.

Calculation of sentence similarity:
The similarity between an input and an example sentence (phrase) is calculated as the sum of the distances between words of the two sentences (phrases). The distance between two words is obtained from a tree-structured thesaurus. Since spontaneous speech has no limit on its vocabulary and incomplete pronunciations as the result of many types of linguistic phenomena, transcriptions of spontaneous dialogues have unknown (out of vocabulary) words. In this processing, robustness to unknown words is achieved by giving a fixed distance for the distance between an unknown word and a word.

Automatic optimization of sentence alignment:

In order to calculate the distance between two sentences, existing EBA processing typically extracts their semantic structures (e.g., case structure) manually. To perform this processing automatically, we apply a dynamic programming technique (DP matching) to obtain optimal alignment between the input and example sentences.

Since a spoken dialogue is transcribed into a single sentence if the partner does not respond, transcriptions have many long sentences composed of short phrases which have little dependence of local context on each other.

For example, the following sentence has four phrases (a phrase is a filler phrase containing no information). "daisan ninnsiki no satou desu ga heya no yoyaku wo sitai n desu ga eeto desu ne getuyoubi no gogo" [I am Satou from the Third Recognition Group, I would like to reserve a room, well, afternoon on Monday].

Sentences of this type are difficult to process with EBA since too many varieties of sentences can be created by the combination of sentences. The lack of similar example sentences may be a serious problem for sentences of this type. To process them, another alignment technique calculates the distance between an input sentence and multiple example sentences. The distance between a sentence and multiple sentences is calculted based on sentence-synchronous continuous dynamic programming technique (continuous DP matching).

Section 4 compares two different alignment techniques.

4 Experiments with EBA Semantic Information Extractor

In a previous EBA experiment that used handmade sentences containing colloquial expressions, where all the words were known, the extractor using out EBA method achieved comparable performance in semantic feature extraction to the rule-

Table 1. **Experiment Conditions**

Vocabulary	422 words
Thesaurus	161 categories maximum 4 layers
Total num. of Example sentences	50 dialogues 458 utterances
Total num. of Test sentences	50 dialogues 513 utterances
Total kinds of Unknown words	102 words 0.4 words per sent.
Matching penalty	deletion 3.0, insertion 5.0
Threshold	3.0

Table 2. **Performance of the Extractor**

Matching Method	Precision (%)	Recall (%)
DP	65.0	44.3
Continuous DP	76.8	58.1
Continuous DP with sentence segmentation	81.6	62.2

based approach[4].

In this study, we performed experiments to extract semantic information from transcriptions of spontaneous dialogues with unknown words. Transcriptions of customer' speech in 50 spontaneous Japanese dialogues were used as examples and transcriptions of another 50 dialogues were used as testing data. Conditions for the experiments are shown in Table 1. Experiments were performed with two optimal alignment methods: DP matching or continuous DP matching.

In experiments with continuous DP matching, two types of example corpus were used either simple transcriptions of utterances or transcriptions of partial utterances that were simple sentences into which utterances were segmented. The aim of this segmentation was to upgrade the example corpus by enlarging the number of corpus sentences by segmentation.

The performance of the extractor was evaluated by two sets of criteria: *precision* and *recall*. *Precision* is the number of semantic features the extractor got correct divided by the number of semantic features the it extracted. *Recall* is the number of semantic features the extractor got correct diveded by the number of possible correct semantic features. Extraction performances in several conditions are shown in Table 2.

The best were a 81.6% precision rate and 62.2% recall rate in semantic features extraction.

The performance with continuous DP matching was 11.8% better in precision and 13.8% better in recall than with DP matching. The difference in performance between DP matching and continuous DP matching mainly resulted from the insufficiency of the example courpus.

Segmenting the utterances improved the performance, but it still remained smaller than the performance achieved by continuous DP matching.

Further improvement should be possible by enlarging the varieties of the example corpus, which will also reduce the performance difference between DP matching and continuous DP matching.

5 Conclusion

This semantic information extractor performs EBA-type processing to extract semantic information essential for the domain of the dialogue without rule-based complex analyses. Extraction is robust against unknown words and many types of ill-formed sentences that contain linguistic phenomena of spontaneous dialogues according to results of evaluation experiments with spoken dialogues about conference scheduling domain.

Acknowledgements

The authors are grateful to Dr. Megumi Kameyama at SRI International and Isao Arima at NTT DATA for their support.

References

[1] M. Mameyama, I. Arima, "A Minimalist Approach to Information Extraction from Spoken Dialogues," *Proc. of ISSD-93*, pp. 137-140, 1993.

[2] M. Nagao, "A Framework of Mechanical Translation between Japanese and English by Analogy Principle" *Artificial and Human Intelligence*, edited by A. Elithorn and R. Banerji, pp. 173-180, 1984.

[3] E. Sumida, "Example-Based Approach in NLP," *SIG-SLUD-9204*, pp. 39-44, 1993. (in Japanese)

[4] O. Shirotsuka, K. Murakami, "A Study of Case-Based Approach to Semantic Feature Extraction from Sentences Using Two-stage DP Matching Processing: Sentence and Phrase," *Proc. of 1994 Spring Meeting of ASJ*, Vol 1, pp. 51-52, 1994. (in Japanese)