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

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

AN ANALYSIS OF JAPANESE SENTENCES IN SPOKEN DIALOGUE AND ITS APPLICATION TO COMMUNICATIVE INTENTION RECOGNITION

Tadahiko Kumamoto, Akira Ito, and Tsuyoshi Ebina

Kansai Advanced Research Center,
Communications Research Laboratory, Ministry of Posts and Telecommunications
Kobe-shi, Hyogo 651-24, Japan

1. INTRODUCTION

The authors are developing a computer-based consultant system. A user can verbally ask this system for help when problems occur. The system understands the user's utterances and then makes appropriate responses. Our target language is spoken Japanese.

To develop such a consultant system, recognizing communicative intention (CI) from a user's utterance is essential. CI recognition consists of two parts, i.e., speech and natural-language processing. In this paper, the latter is focused on and we propose a method of constructing a CI description from a sentence transcribed from the user's utterance [1].

Our approach is based on an analysis of actual user utterances. These utterances were collected in XMH (X-window-based electronic mail handling program) [2] usage experiments. XMH has a visual interface operated using a mouse and a keyboard. Using this XMH, 16 novice XMH users performed a given task, e.g., arranging a tea party. They asked a human consultant for help when they were in trouble and the consultant helped them solve the problem. Based on our analysis of these user utterances, a description form of CI is defined and the CI recognition method is developed by which a CI description is constructed from a sentence.

2. COLLECTING USER UTTERANCES

Actual user utterances were collected in XMH usage experiments (Fig. 1). The user and the consultant were each seated in sound-isolation booths. The user was asked to perform several tasks using XMH. The user asked the consultant's support through a microphone when he/she was in trouble. The consultant listened to him/her through headphones and typed responses on a computer terminal. The responses appeared on the user's display. The experimental subjects were 16 undergraduate students without computer experience.

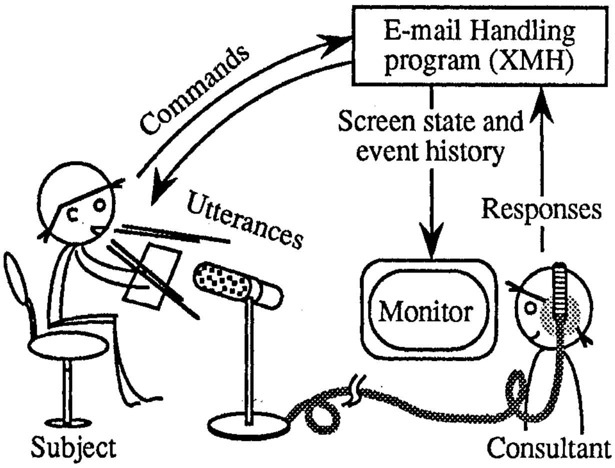

Fig. 1 XMH usage experiment setup.

The collected user utterances were transcribed and a linguistic corpus was constructed. Next, transcription errors and linguistic phenomena such as interjections, repairs, hesitations and paraphrases, were carefully removed. Automatic correction of such linguistic phenomena is one of our future objectives. 24 short sentence fragments of these utterances also were removed. As the results of these modifications, the corpus of 16 data sets was obtained, where each data set corresponds to a user. The corpus consists of 232 simple sentences and 243 complex ones. The 16 data sets were divided into two groups: 10 data sets (298 sentences) and 6 data sets (177 sentences). The former data sets were used for analysis and the latter data sets were used for evaluation of our CI recognition method.

3. DESCRIPTION FORM OF CI

The 298 sentences were analyzed and a description form of CI was defined.

Since an utterance involves propositional information and modality information [3], CI for the utterance should be described in the form:

(CI-type ⟨Proposition⟩),

where CI-type and ⟨Proposition⟩ represent the modality information and the propositional information of the utterance, respectively.

3.1 Defining CI-type

Generally, all kinds of information except for propositional information should be regarded as modality information. However, only the CI type of modality information is used for our consultant system since other kinds of information (nuance, user's emotion, etc.) are not yet technically available. Hence, based on our analysis of the sentences, CI was classified into 12 CI-types (Table 1). Table 1 shows also the sample sentences picked out of our corpus. The parenthesized sentences are the English versions of the sample sentences.

The "ask-wh," "ask-if," "ask-how," and "ask-about" are CI-types for queries asking for an attribute value, a truth value, a procedure and a concept, respectively. The "OK" and "NO" are subtypes for positive and negative queries respectively concerning a task plan. The "EQ" is a subtype for a query concerning the XMH state.

3.2 Representing ⟨Proposition⟩

A frame representation is adopted to represent propositional information of an utterance, which is described in the form: ⟨Proposition⟩.

For example, the propositional information of the sentence "終了を押せばいいんですね (It is alright to push the "Close Window," isn't it?)" is described in the action frame ⟨PushCommandButton #X139EG76⟩ representing the action of pushing the "終了 (Close Window)" button.

4. CONSTRUCTING A CI DESCRIPTION

Based on our analysis of the 298 sentences, a method is developed of recognizing CI, i.e., a method of constructing a CI description.

The flow of CI recognition is shown in Fig. 2. Each stage of the flow is briefly explained.

Table 1. CI-types and sample sentences.

CI-types	Sample Sentences
ask-wh:OK	何を書いたらいいんですか (What should I write here?)
ask-wh	住所忘れました (I forgot his address.)
ask-if:OK	終了を押せばいいんですね (It is alright to push the "Close Window," isn't it?)
ask-if:EQ	メッセージはこれだけでいいんですか (Is this enough for the message?)
ask-if:NO	これ押さなくちゃ駄目なんですか (Do I have to push this?)
ask-if	これで移動できてるんですか (Have I moved it successfully?)
ask-how	移すのはどうするんですか (How do I move it?)
ask-about	スクロールって何ですか (What is "Scroll"?)
have-belief	移動できてないみたいなんですけど (It seems not to have been moved.)
have-goal	ロイの次にカーソルを持っていきたい (I want to move the cursor onto the next of "Roy.")
d-start	すみません (Excuse me.)
d-end	わかりました (I got.)

Morphological analysis and feature extraction stages:

JUMAN [4] was adopted as the morphological analyzer for Japanese sentences, which was provided from Nagao Lab., Kyoto University. JUMAN decomposes an input sentence into several appropriate morphemes and adds information such as: the original form, the "part of speech" name and the conjugation name of the morpheme. Note that our grammatical terminology follows Masuoka-Takubo grammar [5].

A feature vector for the CI-type determination and one for frame generation are separately extracted from the output of JUMAN. Table 2 shows the results of both stages for the input sentence "終了を押せばいいんですね (It is alright to push the "Close Window," isn't it?)."

A feature vector for CI-type determination consists of four elements: semantic words, "part of speech" names, conjugation names and original forms. A "semantic word" is a symbol representing a group of words which gives similar influence on CI-type determination. Semantic words were

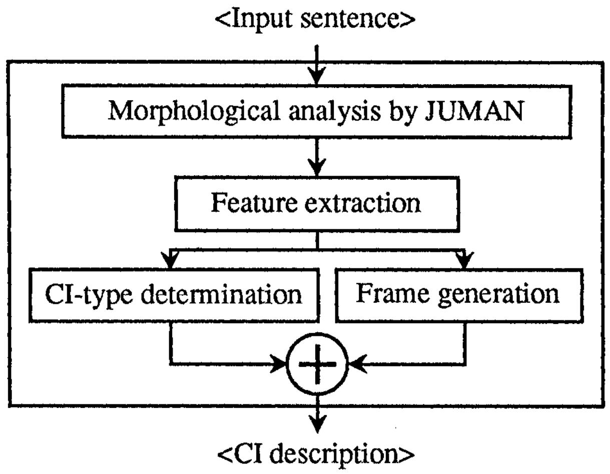

Fig. 2 Flow of CI recognition.

defined based on the analysis of the 298 sentences, some of which are shown in Table 3. A feature vector for frame generation consists of two elements: original forms and "part of speech" names. The values of these elements are extracted mainly from nouns, verbs, adjectives, and postpositional words representing case.

CI-type determination stage:

CI-type is determined by matching a feature vector for CI-type determination with the CI-type determination table (a set of rules). Some rules are shown in Table 4.

The rules were obtained by rearranging all the feature vectors extracted from the 298 sentences for analysis in the following way:

(1) The priority was attached to each semantic word considering its influence on the CI-type determination.

(2) The following operations were done from the semantic word with a higher priority.

(2-1) The sentences from which the feature vectors including the target semantic word were extracted were collected. Suppose that N kinds of CI-types were observed from the collected sentences.

(2-2) For the feature vectors extracted from the collected sentences, the following operation was repeated for $N-1$ times. A feature value or a sequence of feature values was searched for from the feature vectors, which was extracted from all the sentences with the same CI-type but was not extracted from any of the sentences with other CI-types. A list of the searched feature value or sequence of feature values and its CI-types was added as a rule to the table.

Table 2. Outputs of morphological analysis and feature extraction stages.

Feature vector for CI-type determination	Morpheme	Feature vector for frame generation
noun	終了	Close Window: noun
	を	wo: postpositional word representing case
performative-predicate, conditional-form	押せば	push: verb
OK-predicate	いい	
	んです	
pp-word-interrogation	ね	
end-of-sentence		

Table 3. A partial list of semantic words.

Semantic word	Examples
dialogue-start	すみません (Excuse me)
dialogue-end	わかりました (I got)
how-phrase	どう (how)
instruction	教える (Teach me)
want	欲しい (want)
interrogative	誰 (who), 何処 (where)
what	何 (what)
OK-predicate	いい (alright)
pp-word-interrogation	か, よね (postpositional words representing interrogation)

(2-3) The remaining CI-type was given to the sentences to which any of $N-1$ rules obtained in (2-2) didn't applied.

A description form of the table is shown in extended Backus form.

⟨rule⟩ ::= (⟨patterns⟩ ⟨CI-type⟩)
⟨patterns⟩ ::= (⟨pattern⟩ ⟨pattern⟩*)
⟨pattern⟩ ::= (⟨feature value⟩ ⟨feature value⟩*)
⟨feature value⟩ ::= ⟨semantic word⟩ | ⟨"part of speech" name⟩ | ⟨conjugation name⟩ | ⟨original form⟩

How to match a feature vector for CI-type determination with the rules in the CI-type determination table is shown . The following matching is done from top to bottom of the rules in the table.

Table 4. Partial rules in the CI-type determination table.

Patterns	CI-types
(((dialogue-start))	d-start)
(((dialogue-end))	d-end)
(((how-phrase))	ask-how)
(((instruction))	ask-wh)
(((want))	have-goal)
(((interrogative)(OK-predicate))	ask-wh:OK)
(((interrogative))	ask-wh)

Table 5. Accuracy of CI-type determination

	Accuracy
298 learned sentences	100.0%
177 unlearned sentences	89.3%

[Matching a feature vector with ⟨rule⟩]
If the feature values in the feature vector are successfully matched with the ⟨patterns⟩ of the ⟨rule⟩, the corresponding CI-type is determined and the matching is terminated.
[Matching feature values with ⟨patterns⟩]
The matching proceeds from the end of the sentence forward. If a partial sequence of the feature values is the same as the first ⟨pattern⟩, the matching succeeds. If so, the feature values left of the currently matched part are matched with the next ⟨pattern⟩. If the matchings for all the ⟨pattern⟩ succeed, the matching of the feature values and the ⟨patterns⟩ succeeds.

Frame generation stage:

Several frames are generated from a feature vector for frame generation and are semantically bound to each other whenever possible.

For example, the action frame described in **3.2** is generated from the feature vector ("終了 (Close Window): noun" "を (wo): postpositional word representing case" "押す (push): verb").

A CI description is an output combination of the CI-type determination and frame generation stages. For example, the CI description for the input "終了を押せばいいんですね (It is alright to push the "Close Window," isn't it?)" is (ask-if:OK ⟨PushCommandButton #X139EG76⟩).

5. EVALUATION

The CI-type determination method is evaluated using the 177 sentences for evaluation.

First, the CI-types are manually determined for all 475 sentences. Next, the CI-type determination table is made from the 298 sentences for analysis as described in Section **4**. The table has 29 rules, some of which are given in Table 4. Using the table, we made the CI-type determination experiments as follows: Each input sentence is morphologically analyzed, next the feature vector for CI-type determination is extracted, and then the CI-type is determined. If the CI-type determined in this way is equal to the manually predetermined CI-type, the CI-type determination is correct. Otherwise, it is wrong. The results of the experiments are shown in Table 5.

6. CONCLUSION

Actual user utterances were collected in XMH usage experiments, transcribed, and a linguistic corpus was constructed from the transcribed ones. Based on our analysis of the 298 sentences, a description form of CI was defined and a method of constructing a CI description was developed. Using the remaining 177 sentences, the method of determining the CI-type was evaluated. As a result, an accuracy of 100% for the 298 learned sentences and an accuracy of 89.3% for the 177 unlearned sentences were obtained.

REFERENCES

[1] Kumamoto, T., Ito, A., and Ebina, T.: "Recognizing User Communicative Intention in a Dialogue-Based Consultant System — A Statistical Approach Based on the Analysis of Spoken Japanese Sentences —," *Trans. IEICE*, J77-D-II, 6 (Jun. 1994).

[2] J. Peek: "MH and XMH – E-mail for Users and Programmers –", O'Reilly & Associates, Inc. (1991).

[3] Masuoka, T.: "Modarithi No Bunpou," Kuroshio shuppan, Tokyo (1991).

[4] Myoki, Y., Matsumoto, Y., and Nagao, M.: "User Customizable Japanese Dictionary System and Morphological Analyzer," *the 42nd annual meetings of ISPJ*, 1C-9, pp. 3-17 - 3-18 (1991).

[5] Masuoka, T. and Takubo, Y.: "Kiso Nihongo Bunpou," Kuroshio shuppan, Tokyo (1989).