SPEECH UNDERSTANDING BASED ON INTEGRATING CONCEPTS BY CONCEPTUAL DEPENDENCY

Akito Nagai and Yasushi Ishikawa

Human Media Technology Dept.
Information Technology R&D Center
MITSUBISHI Electric Corporation
5–1–1, Ofuna, Kamakura, Kanagawa 247, Japan
Tel. +81 467 41 2077, FAX: +81 467 41 2136, E-mail: nagai@media.isl.melco.co.jp

ABSTRACT

We have proposed a concept-driven semantic interpretation method for a spoken dialogue system that robustly understands various expressions uttered by a naive user. The method is now being improved for practical application. Domain knowledge is important for this improvement. The system must also have portability. This paper discusses the generalization of the semantic interpretation method, and proposes a method that integrates concepts using general linguistic knowledge of conceptual dependency. Speech understanding for various utterances about Kamakura sightseeing with a 1000-word vocabulary was empirically evaluated. The results show that this method can achieve a satisfactory understanding rate.

1. INTRODUCTION

Our goal is to develop a spoken dialogue system that robustly understands various expressions uttered by a naive user in the real world. We have already proposed a concept-driven semantic interpretation method [1] for spontaneous speech understanding.

Two issues from our previous work [2], however, must be solved before our method can be sufficiently improved for practical application. The first is that its semantic constraint was weak so as to attain linguistic robustness. The second is that linguistic knowledge became domain specific and the system lost portability when it was designed in enough detail to attain more accurate performance.

To solve these problems, we introduce general linguistic knowledge of conceptual dependency to attain a general semantic constraint, and we separate domain knowledge and use it as a task-filter that selects a suitable understanding result after general meaning hypotheses are generated.

This paper discusses issues the concerning generalization of constraint knowledge by introducing conceptual dependency, and proposes a new algorithm that integrates concepts based on conceptual dependency. It also discusses how to utilize domain knowledge, and reports the system performance measured by empirical evaluation.

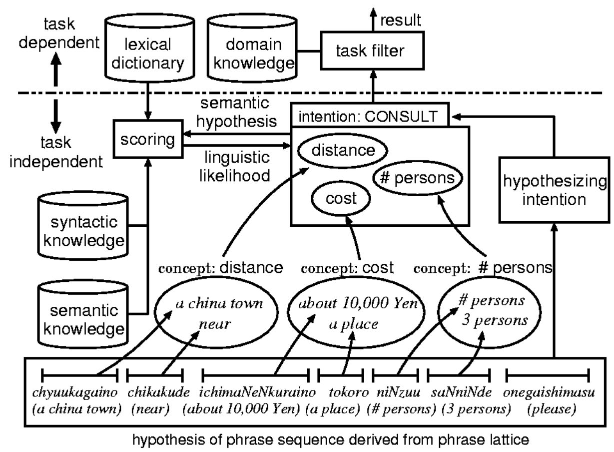

Figure 1: Principle of concept-driven semantic interpretation.

2. KNOWLEDGE GENERALIZATION

Figure 1 illustrates the principle of the concept-driven semantic interpretation method, which is a framework for obtaining both linguistic robustness and constraint. Our approach is based on the idea that a semantic item represented by a partial expression can be a unit of semantic interpretation. We call this unit a *concept*. We consider that; (1) a *concept* is represented by phrases that are continuously uttered as part of a sentence, (2) a sentence is regarded as a sequence of *concepts*, and (3) a user utters *concepts* with an *intention*.

Our earlier method used a semantic frame to represent a concept. First, concepts were detected from a phrase lattice. In detecting concepts, slots were filled by phrase candidates which could be concatenated in the phrase lattice. Second, concept hypotheses were combined using meaning frames. All meaning hypotheses for an entire sentence were generated as meaning frames having slots filled with concept hypotheses.

This method had the following problems regarding linguistic constraint. Senseless combinations of phrases were accepted as concept hypotheses because semantic dependency between phrases was not defined. When we strengthened the linguistic constraint to reduce this problem, the definition of concepts became domain specific. And, concepts that had no relation were accepted as a meaning hypothesis owing to the lack of semantic dependency between concepts.

Therefore, knowledge sources of linguistic constraint are generalized by the following steps;

- domain specific knowledge is separated from the general linguistic knowledge of concepts,
- semantic relations of concepts are defined by conceptual dependency, which represents relations between two concepts, and
- the general form for representing the meaning of an utterance is defined.

2.1. Definition of Concept

To avoid arbitrary decisions in defining concepts, general concepts are objectively selected from a common concept thesaurus. They are classified into three categories; Things (nouns), Actions (verbs) and Attributes (adjectives and adverbs). A hierarchy of concepts is not designed because this may become domain dependent. An example of classified concepts is given in Table 1.

Table 1: An example of classified concepts (partly).

Things: object, goods, conveyance, facility, human, view, nature, food, state, document, symbol, cost, location, area, distance, time, date, etc.
Action: exist, transfer, operate, do, add, result, appear, vary, search, promise, think, trade, perceive, eat, etc.
Attribute: space, shape, hearing, value, useful, state, relation, degree, etc.

2.2. Conceptual Dependency

A semantic relation between two concepts is described by the following form;

$$< conceptA > \Longleftarrow \text{(dependency)} < conceptB >$$

For example, an utterance of *"I'd like to find a hotel."* is represented by a relation of "$< facility >$ $\Longleftarrow$(OBJECT) $< search >$". The relations are (1) cases such as AGENT or OBJECT, (2) attributive modifier (MODIFY), and (3) adverbial modifier (MANNER). Table 2 lists examples of conceptual dependency.

Table 2: Examples of labels for conceptual dependency.

AGENT, OBJECT, IMPLEMENT, SOURCE, GOAL, PLACE, PLACE-FROM, PLACE-TO, TIME, TIME-FROM, TIME-TO, MODIFY, MANNER

2.3. Definition of an Utterance

Our previous representation of an utterance (meaning form) comprised an *intention* and a set of *concepts*. The intention types were defined, for example, as *WH-inquiry (where, when, how, etc.)*, *Yes/No-inquiry*, *reservation*, *change*, *cancel*, and *consultation* for the *Hotel reservation* task. This definition, however, can become domain dependent because it mixes domain-independent intentions with intentions of verbs.

Thus, we have separated domain-independent intentions (*modality*) from those of verbs (*predicative concepts*), and have re-defined meaning form. The new meaning form has registers for an intention, which is represented by a pair of a predicative concept and its modality, and concepts, which are integrated into the predicative concept. The meaning forms are defined for all predicative concepts.

3. UNDERSTANDING PROCESS

On the basis of these general knowledge resources, we have developed a general mechanism to relate concepts and integrate them into the meaning of an utterance. As shown in Figure 2, this algorithm is composed of three processes: to generate original concept hypotheses from a phrase lattice; to integrate them into partial semantic structures of concepts from an original concept lattice; and to construct a meaning representation of an utterance.

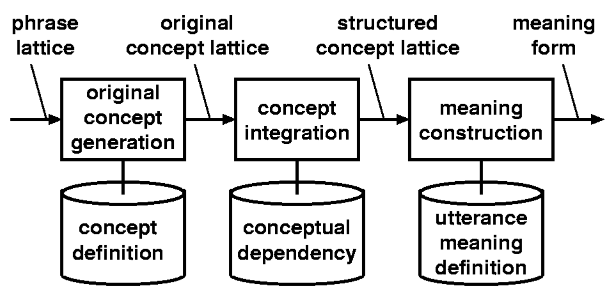

Figure 2: Total process flow.

3.1. Concept Representation

Figure 3 shows a representation of a concept. A concept is represented by a semantic form comprising registers for a name of the concept, other concepts related to the concept, an attribute of the concept (e.g. linguistic properties such as cases and modification), and a recognized phrase.

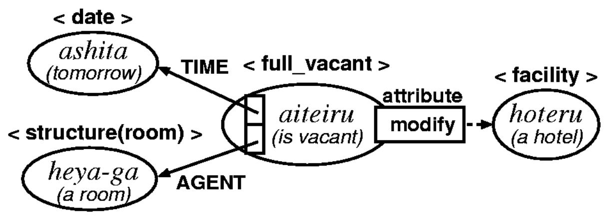

Figure 3: Representation of a concept.

3.2. Integrating Concepts

In the first stage, each phrase candidate is given the name of a concept if a phrase lattice is recognized, then the attributes of the concept are decided by the lexical features of the phrase candidate. These concepts are called original concepts.

Next, two neighboring concepts are combined by conceptual dependency, and are registered to a new concept hypothesis created as shown in Figure 4. A name and attributes of the new hypothesis are given, which are same as them of a main concept (concept B) of the combined two concepts. This integration is repeated between original concepts and new concepts until each concept is integrated. In this way, a lattice of conceptual dependency is generated.

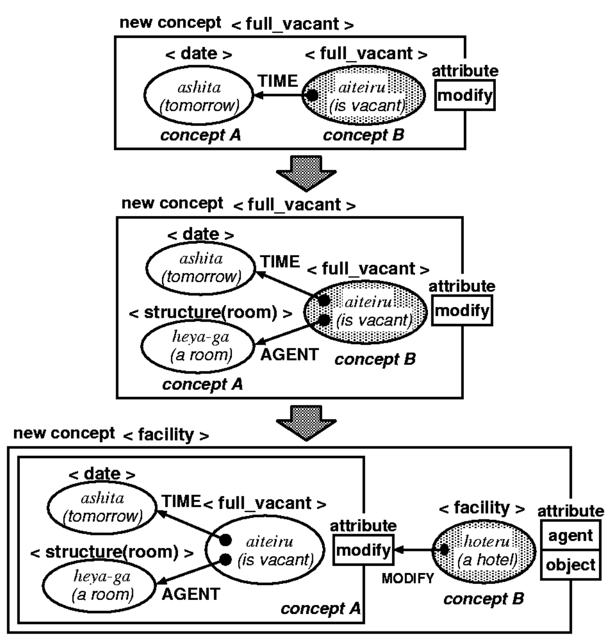

Figure 4: Integration of concepts.

3.3. Constructing Meaning of an Utterance

Detected concepts are combined by the meaning form using the semantic constraint, which is based on conceptual dependency of the predicative concept and its modality. If a concept corresponding to the predicative concept is found, it is registered to the form; otherwise suitable meaning forms are hypothesized for an utterance where a predicative concept is omitted. Finally, integrated meaning hypotheses are accepted.

This algorithm uses a two-stage efficient search method [2] to combine the semantic interpretation with speech recognition (Figure 5). This search method generates initial meaning hypotheses that allow the deletion of concepts. These hypotheses are repaired by re-searching for missing concepts using prediction knowledge associated with the initial meaning hypotheses.

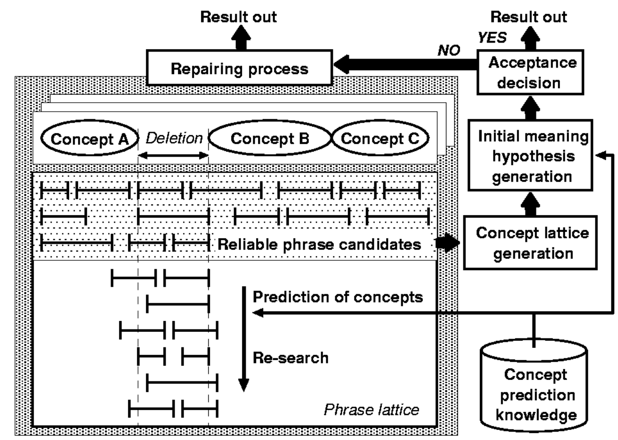

Figure 5: Principle of two-stage search method.

4. EVALUATION

Experiments of speech understanding for various 1000-word-vocabulary utterances about Kamakura sightseeing have been performed.

4.1. Discussion 1

Fifty sentences were made by one subject instructed to make conversational sentences with no limitation of sentential expressions. Three men then uttered these sentences in a soundproof room. Phrase spotting used intra-phrase networks with a 970-word vocabulary and composed of speaker-independent syllable hidden Markov models. The accuracy of the recognized phrase lattices is shown in Table 3.

Table 3: Phrase detection rate (%): 50 utterances, 970-word vocabulary, "N" means a ratio of the number of phrase candidates to the average number of input phrases in an utterance.

lattice size	speaker A	speaker B	speaker C
$N = 1.2$	78.0	79.9	64.5
$N = 5.0$	93.0	98.1	90.7
$N = 10.0$	96.3	98.6	94.9
$N = 20.0$	97.7	99.5	97.7
$N = 40.0$	98.1	100.0	98.6

In the semantic interpretation, 99 concepts, 13 types of conceptual dependency, and 35 meaning forms were used. We used only acoustic likelihood for concept hypotheses, nor linguistic likelihood. In order to evaluate the basic performance of the generalized method, domain knowledge was not used. The standards for judging an answer correct are that an intention (a predicative concept and its modality) is correct, concepts and their boundaries are correct, labels of conceptual dependency are correctly assigned to concept hypotheses, and semantic values of phrase candidates are correctly extracted.

Table 4 lists the understanding rates for three speakers. This shows that high performance can be attained by introducing conceptual dependency between concepts as a general linguistic constraint without domain knowledge. Moreover, we see that the number of concept hypotheses in a concept lattice has been reduced from several hundred in our previous method [2] to 62.0 (ave.). This shows that the conceptual-dependency constraint effectively excludes many senseless hypotheses. We also found that there were mainly two types of understanding error: senseless concepts in terms of domain knowledge and senseless dependency among more than three concepts. Errors in the first case can be removed by utilizing domain knowledge. To exclude the second type of error, we can introduce a constraint to a set of concepts.

Table 4: Understanding rate (%): 50 utterances, 970-word vocabulary.

rank	speaker A	speaker B	speaker C
1	86	86	76
≤ 2	90	88	86
≤ 3	92	90	86
≤ 5	94	94	88
≤ 10	96	98	90

4.2. Discussion 2

We tried to evaluate the generalized method under the condition that the number of false alarms was increased. Spontaneous speech was simulated by instructing subjects to utter the same sentences as those used in the section 4.1. with voluntary filled pauses. Then, phrase lattices of poor quality were made by recognizing the utterances with phrase networks including 23 filled pauses. This meant that many false alarms were included in the phrase lattices. The accuracy of the phrase lattices is listed in Table 5 and the understanding rates in Table 6.

By examining in detail the errors related to speaker A, we found that key-word deletion errors occurred in 7 utterances, false alarms of high likelihood occurred as substitution errors of filled-pauses in 12 utterances, and particle errors occurred in 3 utterances. We found also that substitution errors could be removed with domain knowledge, and that the meaning hypotheses arising from false alarms had semantic validity in terms of general linguistic knowledge. These results convinced us that accuracy could be increased by utilizing domain knowledge.

Table 5: Phrase detection rate (%): 50 utterances, 993-word vocabulary (23 filled pauses), "N" means a ratio of the number of phrase candidates to the average number of input phrases in an utterance.

lattice size	speaker A	speaker B	speaker C
$N = 1.8$	66.2	72.5	60.0
$N = 5.0$	80.4	84.4	75.5
$N = 10.0$	89.5	90.8	85.0
$N = 20.0$	95.4	96.3	88.6
$N = 40.0$	97.3	99.1	93.2

Table 6: Understanding rate (%): simulated spontaneous speech, 50 utterances, 993-word vocabulary (23 filled pauses).

rank	speaker A	speaker B	speaker C
1	44	42	38
≤ 2	60	48	44
≤ 3	66	56	46
≤ 5	68	60	50
≤ 10	72	60	56

5. CONCLUSION

We have proposed a speech understanding method based on integrating concepts related by conceptual dependency. Experimental results of a 1000-word-vocabulary test lead to the conclusion that generalizing the method reduces the number of senseless hypotheses effectively, and that more accurate results can be achieved by utilizing domain knowledge. Future projects will include the introduction of linguistic likelihood for concepts, the construction of a robust speech understanding system for the real world, and the use of domain knowledge for semantic constraint.

6. REFERENCES

[1] Akito Nagai, Yasushi Ishikawa, and Kunio Nakajima, "A Semantic Interpretation Based on Detecting Concepts for Spontaneous Speech Understanding," Proc. ICSLP'94, Yokohama (JAPAN), pp. 95–98, Sep. 1994.

[2] Akito Nagai, Yasushi Ishikawa, and Kunio Nakajima, "Integration of Concept-Driven Semantic Interpretation with Speech Recognition," Proc. ICASSP'96, Atlanta (U.S.A.), pp. 431–434, May 1996.