CONCEPT-DRIVEN SEMANTIC INTERPRETATION FOR ROBUST SPONTANEOUS SPEECH UNDERSTANDING

NAGAI Akito, ISHIKAWA Yasushi and NAKAJIMA Kunio

Computer and Information Systems Laboratory
MITSUBISHI Electric Corporation

ABSTRACT - This paper describes an integration of speech recognition and two-stage semantic interpretation which detects concepts from a phrase lattice and integrates them with an intention to form a meaning hypothesis. In this approach, a concept represented by several phrases is an unit of semantic interpretation, and a sentence is regarded as a sequence of concepts. An island-driven search method exploits syntactic/semantic knowledge to evaluate linguistic likelihood of concept hypotheses. The method has achieved 82% of understanding rate at the first rank (the sixth 94%) by a speech understanding experiment using 50 spoken sentences with various sentential expressions.

INTRODUCTION

To attain linguistic robustness is a very important issue for a spoken language system which enables a naive user to successfully communicate with an intelligent dialogue system (Ward, W. and Young, S. R., 1993), (Zue *et al.*, 1993). In this human-machine communication, a naive user will talk about various semantic items on a task freely in one utterance in the user-initiated manner. These utterances are shown by a large variety of sentential expressions, which are often ill-formed. How does a language model cover such a variety of sentences ?

There is a crucial issue which is closely related with linguistic robustness. How do we exploit linguistic knowledge as a sufficient constraint for attaining better performance of speech recognition ? Syntactic constraint which defines sentential patterns grammatically is contribute to improve speech recognition. But this approach is fragile and no longer robust because it limits sentential expressions and hardly has a sufficient coverage of various utterances of a naive user.

Several recent works have tried to solve these linguistic problems by approaches of relaxing grammatical constraints and partially parsing technique (Stallard, D. and Bobrow, R., 1992), (Seneff, S., 1992), (Baggia, P. and Rullent, C., 1993). These approaches exploit syntactic constraint on an ordinal relation between phrases which occupy a part of a sentence.

In Japanese, however, partial parsing technique is not successful. It analyzes global parts of a sentence as partial trees which consist of phrases ordered grammatically. But spoken Japanese is never represented by even a partial tree because the order of Japanese phrases is considered to be free. Thus, a key issue for attaining linguistic robustness in Japanese spontaneous speech is how to exploit semantic knowledge to represent relations between phrases by semantic-driven processing.

One of the methods of representing a meaning is to use semantic frames which define case relations between phrases from a viewpoint of the predicative usage. In this approach, a hypothesis explosion owing to both word sense ambiguity and many candidates of speech recognition is occurs, if only semantic constraint is used without syntactic constraint. Therefore, a framework to evaluate growing meaning hypotheses based on both syntactic/semantic viewpoints is indispensable in the process of semantic interpretation from a sequences of phrases to a meaning representation.

In our previous work (Nagai *et al.*, 1994), we have proposed a two-stage semantic interpretation method. Our approach is based on an idea that a semantic item represented by a partial expression can be an unit of semantic interpretation. We call this unit a concept. The followings are bases of this method; (1) a concept is represented by several phrases continuously uttered in a part of a sentence, (2) a sentence is regarded as a sequence of concepts, and (3) an user talks about concepts with an intention. We believe that syntactic constraint limited only inside a phrase will be helpful for realizing robustness, and that the explosion will be avoided by treating a concept as a target of semantic constraints.

The purpose of this paper is to give a study of an integration of this semantic interpretation and speech recognition, and to evaluate performance of speech understanding.

PRINCIPLE OF SEMANTIC INTERPRETATION METHOD

The proposed method has two stages of semantic processing (Figure 1). At the first stage, a system hypothesizes an intention of the input as an intention frame. An intention type is associated with verbs and other phrases. At the second stage, the system detects concepts from a phrase lattice using concept frames. Meaning hypotheses of an entire sentence are finally formed by the intention frame filled with concept frames. To reduce ambiguity of the meaning hypotheses, two conditions of existence of a concept and linguistic scoring method are used.

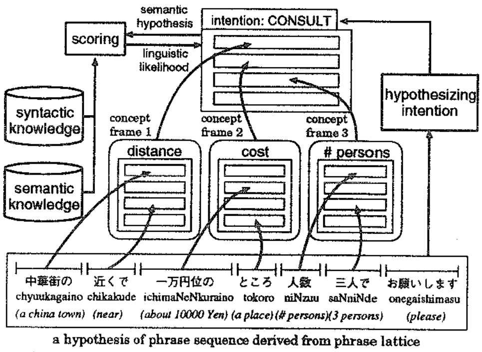

Figure 1: Principle of the semantic interpretation method based on detecting concepts

Detecting concepts

Concepts are defined to represent semantic contents of a target task. Table 1 shows an example of concepts for *Hotel Reservation* task. A concept representation is based on a semantic frame which has several case slots with semantic values and particles as shown in Table 2. Only case slots which are concerned semantically with a concept are defined.

In detecting concepts, slots are filled by phrase candidates which can be concatenated in the phrase lattice, based on examining semantic value and a particle. A phrase candidate which has no particle is examined using only semantic value. This phrase candidate has case-level ambiguity, each case is hypothesized. If a phrase candidate has possibility to be filled to plural slots, each concept hypothesis is produced. To realize robust understanding, a concept hypothesis where all slots are not filled is allowed to exist. Finally, concept hypotheses are combined using intention frames which are associated with sequences of phrase candidates. All meaning hypothesis for an entire sentence are generated as the intention frames which have slots filled with concept hypotheses.

Distance, Place, TakeTime, UseTraffic, Equipment, UseEquipment, Cost, Meal, View, Atmosphere, ReserveDate, ReserveStay, ReserveHotelname, ReserveRoom, ReservePerson, ChangeDate, ChangeStay, ChangePerson, ChangeHotelname, ChangeRoomtype, ChangeRoomNum, Cancel

Table 1: An example of concepts for Hotel Reservation task.

Concept: Distance		
case slot	semantic value	particle
agent	*place	wa, ga
modifier	*place	no
source	*place	kara, yori
goal	*place	made, ni
attribute	*distance	-

note: "-" means that all particle can be supposed.

Table 2: An example of concept representation for *Distance*

DISAMBIGUATION OF CONCEPT HYPOTHESES

To reduce ambiguity of meaning hypotheses, conditions of existence of a concept and linguistic knowledge of scoring it are used. This methods are to filter out senseless concept hypotheses from both syntactic and semantic viewpoints.

Conditions of existence of a concept

Two conditions of existence of a concept are supposed. One is that a concept should have filled slots which are indispensable to forming a gist of the concept. These *indispensable slots* are defined in each concept frame. Another condition is that a concept should occupy a continuous part of a sentence as shown Figure 2. Such a concept hypothesis that has case slots scattered in a sentence is abandoned. This assumes that an user speaks about an semantic item as a chunk of phrases and only once talks about the item in one utterance which includes various other concepts.

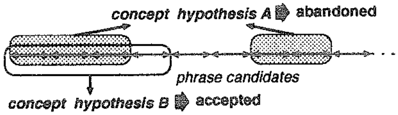

Figure 2: An assumption that a concept should occupy a continuous part of a sentence.

Estimating linguistic likelihood

Syntactic and semantic likelihood of a concept hypothesis is evaluated by a scoring method which considers linguistic dependency between phrases. In this method, penalty rules for a concept hypothesis are used. If a new hypothesis of a concept is produced, it is examined on the bases of all penalty rules. Total score of concept hypotheses are evaluated as linguistic likelihood of a meaning hypothesis. In defining penalty rules, pragmatic knowledge is not used because the proposed method should be as independent of specific task domain as possible. Therefore, penalty rules are defined using general syntactic knowledge of, mainly, particles and modifiers, and shallow semantic knowledge about relation between cases. Several principles of defining penalty rules are shown in Table 3.

Syntactic features	Semantic features
Omission of key particle	Inversion of predicative case and other cases
Inversion of attributive case and substantive case	Predicative case without other cases
Adverbial case without predicative case	Semantic mismatch between phrase candidates
Inadequate conjugation of verbs	Abstract noun without modifiers

Table 3: Principles for defining penalty rules.

EVALUATION

Semantic understanding experiments have been carried out for evaluating disambiguation of concept hypotheses by the scoring method and linguistic robustness of the proposed semantic interpretation method to open test sentences. The input sentences on the *Hotel Reservation* task were made by 10 subjects as character string. The subjects were instructed to make conversational sentences under conditions of Table 4. The sentences of test set A which should be plain were, however, considerably represented by various expressions as a result.

set name	#sent.	conditions to subjects
set A	51	to make as plain sentences as possible
set B	64	to make various sentences freely (no limitations)

Table 4: Two test sets of sentences (*Hotel Reservation*)

Procedure

Test sentences were described as sequences of phonetic symbols. A lattice generator analyzes the test sentence by driving chart parser using phrase grammar of approximately 1000-word vocabulary and outputs all possible phrase candidates as a lattice. Each phrase candidate is given a semantic value looking up a semantic dictionary. Semantic interpreting process decides an intention and selects the intention frame for each phrase sequence (same structure of intention frame has been used in the experiments). The process detects concepts using 22 concept frames, filling slots with phrase candidates and examining conditions of concept existence. Then, all meaning hypotheses are penalized to find the best hypothesis from a viewpoint of linguistic likelihood. Penalty of one is added to a meaning hypothesis every time it has a linguistic feature of penalty rules.

Results and discussion

Two kinds of experiments have been performed. The scoring methods has been evaluated for test set A as reported in Table 5. The candidates stands for meaning hypotheses which have same linguistic likelihood at the first rank. So the number of candidates is ambiguity of meaning hypotheses. The understanding rate stands for a rate of results which include a correct hypothesis in the candidates.

Syntactic penalty rules reduced the ambiguity of 50.8 to 4.1 without missing a correct hypothesis. The rules worked well in disambiguation on syntactic modification like attributive case. Using all rules including semantic penalty rules achieved almost complete disambiguation to 1.1. But a correct hypothesis was missed in the case of conceptual utterance where predicative case and other cases did not coexist (case omission).

Linguistic robustness of the method has been evaluated for test set B. 75.0 % (48 sentences) of sentences were correctly understood as shown in Table 6. 20.3 % (13 sentences) were rejected including linguistic phenomena of a complex sentence, repetition or restart, a chain of same case element and unknown concepts. As for a complex sentence, one of approaches is that concept of a relative clause is registered as attributive case slot of a subjective concept including the antecedent. The utterance including repetition can be understood merging a repeated concept into a concept already hypothesized. If these *No Answer* cases are excluded, the proposed method showed a performance of 94.1% (48 / 51 sentences) with ambiguity of 1.5.

Conditions	#Candidates	Understanding Rate(%)
(1)	50.8	100.0
(1)+(2)	4.1	100.0
(1)+(2)+(3)	1.1	98.0

Table 5: Test set A (48 sentence). Conditions are; (1) conditions of existence of a concept, (2) syntactic scoring rules, (3) semantic scoring rules.

Conditions	#Candidates	Correct	Error	No Answer
(1)+(2)+(3)	1.5	75.0%	4.7%	20.3%

Table 6: Test set B (64 sentences). Conditions are; (1) conditions of existence of a concept, (2) syntactic scoring rules, (3) semantic scoring rules.

INTEGRATION OF SPEECH RECOGNITION

Towards integrating the semantic interpreting and speech recognition, we take an approach of island-driven search method for detecting concept hypotheses. Figure 3 outlines this concept-driven search process. First, speech recognizer based on phrase spotting sends a phrase lattice and pause hypotheses to the semantic interpreter. Then, phrase candidates are selected as seeds for growing concept hypotheses.

Two semantic processes are performed after selecting seeds; concept lattice generation and meaning hypotheses search. In the stage of concept lattice generation, each concept hypothesis is extended in the direction of both forward and backward considering existence of gaps, overlaps and pauses. In the next stage, meaning hypotheses search, concept hypotheses are concatenated to form meaning hypotheses of a whole sentence.

This concept-driven search method exploits syntactic/semantic knowledge as well as acoustic knowledge. The linguistic scoring method evaluates growing concept hypotheses and abandons hopeless hypotheses in both stages of two semantic processes. Acoustic likelihood of a concept hypothesis is defined as an average of phrase likelihood which is normalized by phrase duration. Linguistic likelihood is given to each concept hypothesis using the syntactic/semantic penalty rules.

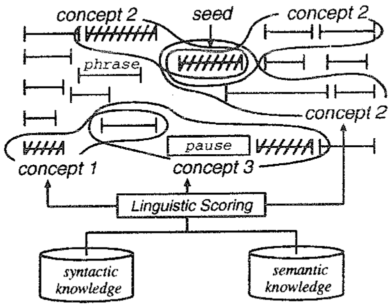

Figure 3: Island-driven search for detecting concept hypotheses

Concept lattice generation

Several criteria of extending concept hypotheses are used as follows;

1. Gaps and overlaps between phrases are permitted, if their length is within permitted limit.
2. Pauses are also permitted between phrases. The speech recognizer detects pauses as labels with start and end frame information. These pause labels are permitted to be connected to phrase candidates considering gaps and overlaps within the permitted limit.
3. Phrase candidates which satisfies two conditions of existence of a concept are connected. Indispensable slots, however, are evaluated after a concept lattice is generated in order to avoid missing seeds, because seeds are not always phrase candidates for the indispensable slots.
4. Both acoustic and linguistic likelihood are estimated and given to a concept hypothesis whenever the hypothesis is extended to integrate a phrase candidate. If acoustic or linguistic likelihood of a concept hypothesis is worse than its thresholds, the hypothesis is abandoned.

Meaning hypotheses search

Meaning hypotheses for a whole sentence are generated by concatenating concept hypotheses in the concept lattice. This search process is performed in a best-first manner based on acoustic likelihood. In connecting concept hypotheses, linguistic likelihood of growing meaning hypotheses is also evaluated and existence of gaps, overlaps and pauses are considered between concept hypotheses with in the permitted limit. Moreover, phrase candidates of verbs which are defined to relate to intention types are included in meaning hypotheses. We can choose a method for ordering accepted meaning hypotheses in these; (1) they are ordered based on acoustic likelihood. The hypotheses having same acoustic likelihood are ordered based on linguistic likelihood. (2) a method contrary to (1), (3) based on a merged score of acoustic/linguistic likelihood. In this paper, we take the method (1).

Results and discussion

A speech understanding experiment was performed on a test set of 50 spoken sentences (48: set A, 2: set B) which were uttered by 1 male speaker. The average number of phrases is 5.8 per a sentence. Intra-phrase grammar which included 375 rules with 356-word vocabulary (330 content words, 26 function words) was converted into phrase networks for phrase spotting recognizer, which used speaker-independent and context-independent phonemic HMMs. Phrase lattices provided for semantic interpretation process include false alarms from 10 to 30 times the average number of input phrases. Detecting rate of correct content words is 100% without considering function words (particles) errors. The permitted limit of gaps and overlaps duration is 13 frames (10 ms frame period).

Table 7 reports the result of semantic understanding experiment. The standards of judgment for correct answer are as follows; (1) sorts of concepts and their boundaries are correctly detected, (2) cases of the concept are correctly assigned to phrase candidates, (3) semantic values are correctly extracted. The result shows that the good performance of 76% at the first rank and 90% in the ninth rank have been achieved.

The understanding errors are mainly caused by following cases; (1) connecting of correct phrase candidates fails because length of inter-phrase gaps exceeds the permitted limit of gap owing to deletion errors of particles and pauses, (2) seeds which grow to correct concept hypotheses are not detected in the seed selection stage and (3) a rank of a correct meaning hypothesis considerably decreases if linguistic likelihood given as an integer differs by one, because it is roughly estimated comparing acoustic likelihood and there are many competitive hypotheses which have same linguistic likelihood. To cope with these errors, (1) speech recognition has to be more improved using, for example, context-dependent precise HMMs, (2) the number of seeds should be sufficient to cover correct concepts, moreover, search strategy considering the seed deletion error is required and (3) scoring and ordering method merging both acoustic and linguistic likelihood will enable more precise evaluation.

rank	understanding rate
1	82 %
~ 2	84 %
~ 3	90 %
~ 4	92 %
~ 6	94 %

Table 7: Results of speech understanding experiment: 50 test sentences, 1 male speaker

CONCLUDING REMARKS

We proposed a two-stage semantic interpretation method for robustly understanding spontaneous speech and described the integration of speech recognition. In this approach, a concept plays important roles; as a robust interpreter of various partial expressions, as a target of semantic constraint and as a basic unit of understanding a whole meaning. The experimental results show that the linguistic scoring method greatly contributes to disambiguation of concept hypotheses and 75% of test sentences made by subjects with no limitation are correctly understood with ambiguity of 1.5.

This semantic interpretation is successfully integrated with speech recognition by island-driven lattice search method and achieves good performance showing 82% of semantic understanding rate at the first rank and 94% in the sixth rank. These results encourage us to realize a spoken language system which robustly understands various utterances of naive users.

Future enhancements will include; (1) scoring method merging acoustic/linguistic likelihood, (2) refining concept-driven search algorithm, (3) improving the interpretation for a complex or compound sentence which is not a subject of this paper, (4) exploiting top-down knowledge for speech recognition and (5) more detail evaluation using spontaneous speech data of naive speakers.

REFERENCES

Baggia, P. and Rullent, C. (1993) *Partial Parsing as Robust Parsing Strategy*, Proc. ICASSP 93, II-123–126.

Nagai, A., Ishikawa, Y. and Nakajima, K. (1994) *A SEMANTIC INTERPRETATION BASED ON DETECTING CONCEPTS FOR SPONTANEOUS SPEECH UNDERSTANDING*, Proc. ICSLP 94, S04-6, 95–98.

Seneff, S. (1992) *A Relaxation Method for Understanding Spontaneous Speech Utterances*, Proc. DARPA Speech and Natural Language Workshop, 299–304.

Stallard, D. and Bobrow, R. (1992) *Fragment Processing in the DELPHI System*, Proc. DARPA Speech and Natural Language Workshop, 305–310.

Ward, W. and Young, S. R. (1993) *Flexible Use of Semantic Constraints in Speech Recognition*, Proc. ICASSP 93, II-49–50.

Zue, V., Seneff, S., Polifroni, J., Phillips, M., Pao, C., Goodine, D., Goddeau, D. and Glass, J. (1993) PEGASUS: *A Spoken Dialogue Interface for On-Line Air Travel Planning*, Proc. ISSD–93 (Waseda Univ., Tokyo, Japan), 157–160.