DIALOG CONTEXT DEPENDENCIES OF UTTERANCES GENERATED FROM CONCEPT REPERESENTATION

Yoichi Yamashita[1] *Keiichi Tajima*[2] *Yasuo Nomura*[2] *Riichiro Mizoguchi*[1]

[1]I.S.I.R., Osaka University, Ibaraki-shi, Osaka, 567 Japan
[2]Faculty of Engineering, Kansai University, Suita-shi, Osaka, 564 Japan

ABSTRACT

In our framework of the speech output interface based on concept-to-speech conversion, sentence generation according to dialog context is a crucial issue to be discussed. Understanding of the diversity of surface sentences and relation between them and dialog context enables such an interface system to generate spoken utterances with high quality. This paper describes two experiments in order to investigate dialog context dependencies of utterances in written dialog. In the first experiment, sentences are twice generated from the same concept representation under two situations; sentences are generated isolatedly or generated by referring the preceding utterances in the dialog. A lot of differences are observed between a set of sentence pairs and classified into 22 categories. The second experiment examined preference of various expressions based on the categories obtained from the first experiment and verified substantiality of each category for sentence generation according to dialog context. Mechanisms for dialog processing in the interface system are also discussed at the last of the paper.

I. INTRODUCTION

One of goals of speech research is to build a general speech interface through which we can communicate with various intelligent performance systems (IPS). Our framework of speech output interface based on the concept-to-speech (CTS) conversion system, named SOCS, is designed to be independent of the IPS and converts concept representation generated by the IPS into synthetic speech[4]. In the CTS conversion, sentence generation is a very important issue as well as speech wave production. The process of sentence generation is often divided into two stages; decision of 'what-to-say' and 'how-to-say'. Importance of interactions between these two stages is pointed out[1], but this scheme of two stages is suitable for separation of general functions of the interface component from the problem solving in the IPS, that is application. Our speech interface is responsible for decision of 'how-to-say', that is generating spoken utterances according to dialog context, after the IPS determined 'what to say' to the user.

Former studies on sentence generation paid attention to the cohesion and the coherence in discourse[2, 3]. In dialog, different expressions for the same meaning are produced with varieties of the surface sentence and the prosodic characteristics according to dialog context. Understanding of the diversity of surface sentences and relation between them and dialog context is indispensable for generating spoken utterances with high quality.

In this paper, we analyze dialog context dependencies of the surface sentence through an experiment of sentence generation in order to find out roles of the dialog manager in the interface for generating sentences. Preference of several expressions which are caused by dialog context is also investigated through another experiment. Some mechanisms are discussed for generating sentences taking account of dialog context in the dialog manager.

II. SPEECH OUTPUT BASED ON CONCEPT-TO-SPEECH CONVERSION

CTS has following advantages over conventional text-to-speech (TTS) conversion;

(1) there is no need for text analysis because CTS generates sentences from concept representation for itself,

(2) some prosodic features, such as emphasis, speaking rate, and so on, can be embedded in concept representation, and

(3) a concept representation can be transformed into various sentences according to dialog context.

The last advantage indicates that CTS plays an important roles for spoken dialog synthesis in human-computer interaction. An IPS on computer does not have to consider dialog context and determines only what to say to the user using concept representation. The speech synthesis based on the CTS system generates spoken utterances taking account of dialog context. CTS can be one of key techniques for human-computer interaction through spoken language.

From these viewpoints, the authors have designed a framework of an interface for speech input and output, shown in Figure 1, which is independent of the IPS, that is an application. This interface system is composed of three major components; the speech understanding component, the CTS component and the dialog manager. The IPS should decide only 'what-to-say' and commit the determination of the surface spoken language expression to the speech interface system. The dialog manager modifies the concept representation according to dialog context. For example, features necessary to prosody generation are added to the concept representation and redundant concepts in

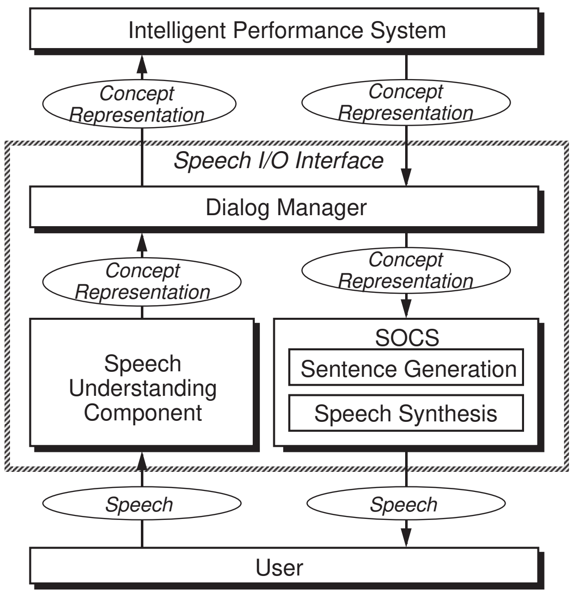

Figure 1. A block diagram of the speech interface.

aru([$actor($and(*kouen*,*kanketsusen*), $place($modify(*kamisuwaeki*,*chikaku*))]).
'exist' 'park' 'geyser' 'Kamisuwa station' 'near'

$is(*kanketsusen*, *utsukushii*).
'beautiful'

Figure 2. Examples of concept representation.

U1: *shukuhaku no yoyaku o totte nainde, dokoka annaishite kuremasuka*
'I have no reservation. May I have some information?'

S1: *hotel to ryokan ga arimasuga, dochira ga iidesuka*
'Which do you want, a hotel or a Japanese style hotel?'

U2: *dokodemo iinode kankoukousu no soba de onegai shimasu*
'I don't care. Near the route of sight seeing, please.'

S2: *susumeru*([$obj(*hoteru osaka*]).
'recommend' 'the Hotel Osaka'

/without context/ *hoteru osaka ga osusume desu*
'I recommend the Hotel Osaka.'

/with context/ *sorenara, hoteru osaka ga osusume desu*
'In that case, I recommend the Hotel Osaka.'

Figure 3. An examples of the conjunction insertion.

the dialog context are removed. Then, the CTS component, named SOCS, generates sentences and converts them into synthetic speech.

The representation scheme describing output messages is an crucial issue for speech output from the IPS. If the messages are expressed in natural language text, modification of the messages according to dialog context becomes difficult in the interface system. The representation form with an appropriate abstraction level is very important for describing messages that are inputs to the interface system for speech output. The authors defines representation based on the case structure and the phrase patterns as concept representation, considering the interaction between the IPS and the CTS component[5]. Figure 2 shows examples of concept representation. Both the message generated by the IPS and the input to SOCS are described with the same concept representation scheme. However, the contents of them may be slightly different, because the dialog manager modifies the concept representation according to dialog context.

III. EXPERIMENTS

3.1. Sentence generation

In general, dialog contexts cause different expressions of the sentence for the same meaning. The diversity of such various sentences must be investigated before the discussion of the dialog management for sentence generation in the general speech interface system. An experiment of sentence generation from concept representation is carried out under two situations; sentences are generated isolatedly or generated by referring the preceding utterances in the dialog. Each situation is called sentence generation without and with dialog context, respectively. A set of concept representations is prepared by referring a scenario arranged from a simulated dialog between a questioner and an answerer. Only answerer's utterances are converted into concept representations by hand. Each representation corresponds to a utterance in the dialog.

First, a subjects is given the concept representations in a random order, that is, without dialog context, and he/she converts them into isolated sentences one by one. Then, he/she is given the same set of concept representations in the same order as in the dialog, that is, with dialog context. The subject generated sentences again. We used two scenarios of dialog and the number of answerer's utterance is 62 and 60 for each scenario. The number of subjects is 3 and 4, respectively.

The differences between two sets of sentences are analyzed and classified into following 22 categories. The observation count for each difference is summarized in Table 1.

(A) Insertion of conjunctions
In the case of the isolated sentence generation without dialog context, the conjunction rarely appears. When dialog context is given, the subjects sometimes insert the appropriate conjunction at the beginning of the sentence in order to clarify the relation to the preceding utterance.

(A-1) The conjunctive word which indicates affirmative connection, such as '*sorenara*(in that case)', '*dewa*(then)', and so on, is inserted after a requirement of the opponent. Figure 3 is this example.

(A-2) When a new topic is introduced after termination of the previous topic, the conjunction such as '*sore-dewa*(then)' are often added.

(A-3) Some utterances are coordinate with the preceding utterance, the coordinate conjunction of '*soshite*(and)' clarifies the relation between two utterances.

(B) Change of the word order
Japanese sentences are free of the word order and the order of case informations for a predicate is not indicated in concept representation. Dialog context moves the given case information forward in the sentence.

(C) Move of the center in answers
Some utterances are composed of several concept representations and the subject often converts them into simple sentences. When they are an answer to the preceding utterance, the representation relating to the center in the answer is generated at the top of the simple sentence.

(D) Clarification of the center in answers
When an answer is composed of several concept representation clauses, the representation which does not describe the center in the answer but modifies the center makes a subordinate clause.

(E) Ellipses
Given concepts are often omitted in the surface sentence under the situation with dialog context even if they are described in the concept representation.

(F) Description of the omitted words
The subjects sometimes compensate omitted words which are not described in the concept representation in order to perform reliable information transfer.

(G) Specification using modifiers
The demonstrative pronouns such as '*sono*(that)' are added to a sentence so that an instance of concept in the

Table 1. Analysis of experiments of sentence generation and preference of each factor.

	category	frequency	preference rate
	Insertion of conjunctions		
A-1	after a requirement	8	88 %
A-2	after topic change	5	85 %
A-3	to clarify the coordinate sentence	21	75 %
B	Change of the word order	23	69 %
C	Move of the center in answers	4	73 %
D	Clarification of the center in answers	3	77 %
E	Ellipses	14	61 %
F	Description of the omitted words	4	90 %
G	Specification using modifiers	5	75 %
	Change/addition of function words		
H-1	'*ga*' → '*wa*'	5	83 %
H-2	'*wa*' → '*ga*'	5	68 %
H-3	to clarify a topic	11	62 %
H-4	to specify similar information	22	90 %
H-5	to clarify comparison	1	80 %
	Change/addition of content words		
I-1	euphemism	2	100 %
I-2	tag question	12	46 %
I-3	idiomatic expression	3	59 %
I-4	confirmation	3	100 %
I-5	additional meanings	15	71 %
I-6	to introduce the assumed topic	14	77 %
J	Insertion of "yes" or "no"	2	65 %
K	insertion of idioms	10	73 %
	Total / Average	192	71 %

(a)	*kyoto de nani o mitai desuka*	(A-2:off, H-3:off)
(b)	*kyoto dewa nani o mitai desuka*	(A-2:off, H-3:on)
(c)	*soredewa mazu, kyoto de nani o mitai desuka*	(A-2:on, H-3:off)
(d)	*soredewa mazu, kyoto dewa nani o mitai desuka*	(A-2:on, H-3:on)
	'Well, what would like to see in Kyoto?'	

Figure 4. An Example of the coices.

preceding utterance is identified as those in the sentence to be generated.

(**H**) Change/addition of function words

Postpositional particles play important roles in a Japanese sentence for denoting a topic and expressing a delicate shade of meaning as well as indicating the case of the content words. Dialog context causes change or addition of propositional particles.

(H-1) The propositional particle '*ga*' denoting the subject in a sentence is replaced by '*wa*' so as to clarify the topic.

(H-2) On the contrary, the particle '*wa*' is changed into '*ga*'.

(H-3) The particle '*wa*' follows other particles denoting the case and clarifies the topic. This is similar to H-1.

(H-4) Other propositional particles, '*nado*(such as)' and '*mo*(too)', are inserted so as to indicate that a content word is the same kind of information as given one.

(H-5) When a utterance contains two comparative concepts, a particle '*wa*' make the comparison clear.

(**I**) Change/addition of content words

Differences of content words are found between the generated sentence without and with dialog context. Besides conjunctions and propositional particles, some words are changed or added for the smooth communication in dialog.

(I-1) The subjects sometimes prefer euphemistic expressions replacing a WH-question by superficial YN-question which asks the existence, such as '$\cdots$ *wa arimasuka*(Are there $\cdots$)'.

(I-2) The subjects change an answer/assertion sentence into a sort of Japanese tag question replacing the end of the sentence by a phrase '*desuga*' that implies a question and wants an answer.

(I-3) A sentence generated according to dialog context has an idiomatic expression fitted to the dialog.

(I-4) An YN-question is changed into a confirmation sentence which requires permission.

(I-5) The subjects compensate words into the sentence according to dialog context. This is similar to F, but is addition of clauses or partial sentences while F is recovery of the case information.

(I-6) An auxiliary verb '*nara*' is used for denoting a assuming topic in the sentence.

(**J**) Insertion of "yes" or "no"

The word explicitly expressing affirmation or negation, that is '*hai*(yes)' or '*iie*(no)', is inserted at the beginning of an answer utterance.

(**K**) insertion of idioms

Some idiomatic phrases and interjections, such as a greeting, '*soudesune*(well)', and so on, can be found in the sentences with dialog context.

3.2. Preference of several expressions

Speech output based on the CTS architecture for human-computer interaction needs sentence generation according to dialog context. Various differences of the sentence between without and with dialog context were observed through the experiment of the sentence generation from concept representation. However, all subjects did not generate the same expression of the sentence. Some differences were given by one subject or observed with a low frequency. It is afraid that the generated sentences may include much fluctuation and that some categories are not substantial for the dialog context dependency of the surface sentence. The dialog manager in the general speech interface, mentioned in Section 2, is responsible for incorporating dialog context into the sentence generation process. It is very important to determine what factors are essential to the sentence generation according to dialog context.

We carried out another experiment in order to investigate the preference of various expressions. A subject is presented with several choices of a sentence along with the dialog context and selects the best one fitted with the context. The choices are prepared by the combination of some categories in Table 1. What categories to be involved for each sentence is determined based on the first experiment of sentence generation. When two or three categories are involved, four or eight choices are presented at once for a sentence, respectively. If contradictory categories (for example, C and E in Table 1) are involved, choices from either of them are removed. 10 subjects selected the best choice for 65 sentences. Each sentence had 3.9 choices in average. Figure 4 shows an example of choices. In this case, categories A-2 and H-3 are involved and four choices are presented to the subject. If he/she selected (b), H-3 is preferred and A-2 is not. It is assumed that each category of sentence variations is independent each other for the preference. Table 1 also summarizes the result of the second experiment. The higher the rate is, the more the subjects preferred. The rate of 50% is neutral. Average preference rate is 71% and many kinds of categories got high scores over 80%. The dialog processing for such preferred categories should be incorporated into sentence generation process in the general speech interface.

IV. DIALOG PROCESSING MECHANISM

The experiment of the sentence generation from concept representation pointed out that dialog context variously changes the surface sentence. The second experiment indicated that most of the categories of the differences between without and with dialog context are important to produce a sentence expression appropriate to dialog context. It is necessary to incorporate dialog processing for sentence generation with the speech interface system based on CTS and to adapt input of concept representation to the dialog context. The differences summarized in Table 1 were caused by the presence of the dialog context. However, the subjects sometimes generated sentences using the domain knowledge and the user model as well as the dialog context. It is difficult to deal with such information in the current design of our interface system because it is independent of the IPS and does not possess the user model. The knowledge transfer from the IPS to the interface and the user model will be a future work. The mechanisms of dialog processing based on general knowledge on dialog are discussed for two categories in this section.

4.1. Conjunctions after a requirement

Several kinds of the conjunction or the conjunctive phrase are inserted at the beginning of an answer after a requirement utterance, which is a question or a request, as mentioned as A-1 in section 3.1. However, this is always not the case. It is necessary to identify the dialog context, that is the situation of inserting such conjunctions. A typical example of the inserted conjunction '*sorenara*(then)' means that a given concept is accepted as a condition. This is modeled by introducing utterance type as follows;

> Insert a conjunction '*sorenara*' if the utterance (*S-Res*) to be generated is <Inform-Value>, the opponent's utterance (*U-Req*) paired with *S-Res* is <Ask-Value>, and there is another <Inform-Value> utterance nested between the utterance pair of *U-Req* and *S-Res*.

Here, <Inform-Value> and <Ask-Value> are utterance types which represent a utterance telling values of the object and a utterance asking something typically in WH-questions, respectively. An example of inserting the conjunction shown in Figure 3 is this case. The utterance *U1* and *S2* is <Ask-Value> and <Inform-Value>, respectively, and forms a utterance pair. The nested *U2* is <Inform-Value> for telling supplementary information.

4.2. Change of word order

Japanese sentences are free of the constraint of the word corder. We can understand the meaning of a sentence even if the positions of the case for a predicate are exchanged each other in the sentence. However, the word order is very important to naturalness of the sentence. In the experiment of sentence generation, the subjects changed the sentences into those with the appropriate word order to the dialog context.

When we communicate using natural language, the information known each other is confirmed first and something new which we want to tell is described later. Such a manner of building a utterance or a sentence makes communication smooth and facilitates the listener's understanding. This process of generation is easily controlled by using stacks. The dialog manager must store given information in the stack and refers the stack before the sentence generation. The dialog manager communicates messages of input or output with other components using concept representation in our framework. This scheme enables to manage given information in the concept level. The given information should be flushed after a new topic is introduced and the subject of the dialog is changed. The topic understanding is also incorporated with the dialog management.

V. CONCLUSIONS

The experiment of sentence generation under two situations without and with dialog context verified that the most appropriate sentence expression is much dependent on dialog context. Differences of the generated sentences are classified into 22 categories. Some categories, such as change of function words, idiomatic expression, and so on, are of course dependent on Japanese. We believe that many categories are the case for other languages. Another experiment of preference to expressions of the surface sentence made it clear that most of the differences observed in the first experiment are substantial for utterance generation in dialog. Properly speaking, we should investigate the dialog context dependencies with spoken utterances in order to realize the human-computer interaction though spoken language. The results of the experiment based on written sentences are very useful for the dialog management for spoken dialog.

REFERENCES

[1] Appelt, D.E.: "Planning English Sentences", Cambridge University Press (1985).

[2] McDonald, D.D.: "Natural language generation", In Shapiro, S. (ed.), *Encyclopedia of Artificial Intelligence: Second Edition*, pp.983–997, John Wiley & Sons, Inc. (1991).

[3] Mykowiecka, A.: "Natural-Language Generation – an Overview", *Int. J. Man-Machine Studies*, 34, pp.497–511 (1991).

[4] Yamashita, Y. and Mizoguchi, R.: "SOCS: A Speech Output System from Concept Representation", *Proc. of ICASSP '92*, 2, pp.69–72 (1992).

[5] Yamashita, Y., Mizutani, N., Kakusho, O. and Mizoguchi, R., "Speech Synthesis from Concept Representation in General Speech Output Interface", *Trans. of IEICE*, J76-DII, 3, pp.415–426, 1993 (in Japanese). (also to be appeared in *Systems and Computers in Japan*)