Generating Spontaneous Elliptical Utterance

Yuji Sagawa *Wataru Sugimoto* *Noboru Ohnishi*

Department of Information Engineering,
School of Engineering, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-01 Japan
sagawa@nuie.nagoya-u.ac.jp

ABSTRACT

To generate spontaneous utterances, it is important to omit some items from them appropriately. Previously, generation of elliptical utterances is discussed only from a view of whether an item in question is known or unknown in the context of dialogue. We point out that some items that are already known should not be omitted in some situation. For example, omitting too many items from utterances makes a hearer to feel that the speaker is impolite.

In this paper, we propose some strategies to generate appropriate elliptical English utterances, given a politeness level. Our system can translate utterances with no ellipses into appropriate elliptical utterances using the strategies. The strategies are obtained by analysis of some dialogues. But we show that they are applicable to other dialogues.

1 INTRODUCTION

In a spontaneous dialogue, a human speaker often use elliptical expressions to reduce communicative efforts of the speaker and the hearer. If someone always utters without ellipsis, a hearer may feel redundant and get irritated. So, to determine which item to be omitted from utterance is an important problem in any natural language generation system.

A main purpose of ellipsis is to reduce hearer's effort by not referring to an easily identified item from the context of dialogue. Indeed, a main criterion for previous discussions about generating elliptical utterances is that it is better to omit as many items as possible if the items can be identified from the context of dialogue.

We think this is not enough. If one can see dialogue only as a tool for conveying information or intention of the speaker, it is enough. But in communication, we meta-communicate various information such as politeness or attention of the speaker besides the content of utterances[1][2].

Our main claim is that whether an item is omitted or not is determined not only by whether it is known or unknown but also how polite the speaker wants to be. To be very polite, the speaker often refers to an item that is already known.

In this paper, we propose some strategies to generate appropriate elliptical English utterances, given a politeness level.

Then, we show the outline of a system which generates elliptical utterances. The system can translate utterances with no ellipsis into appropriate elliptical version using the strategies. By using the system as a final module of a natural language generation system, one can get appropriate elliptical utterances.

Finally, we show that they are applicable to another dialogues from ones used for our analysis to show how general our strategies are.

2 GENERATING ELLIPTICAL UTTERANCES

Some works have been done on generating elliptical Japanese utterances [3][4]. In English, few works have been done on generating elliptical utterances because pronouns are preferred to ellipsis in English. But informal dialogues contain many ellipses also in English.

A main criterion for previous discussions about generating elliptical utterances is that it is better to omit as

many items as possible if the items can be identified from the context. In other words, because the meaning of an elliptical utterance and that of non-elliptical version of the utterance are the same, a shorter utterance is preferred.

As an example, consider utterances (1) and (2) as replies to an question *"May I change my seat to the center near the aisle?"*.

(1) *Yes, of course you can change it.*

(2) *Yes.*

In (2), the phrase *"of course you can change it"* is omitted. Both utterances have the same content, say the fact that the speaker of the question can change his seat, because even in (2) what the speaker approves is obvious from the context.

Though two utterances has the same content, We think they are not quite the same.

The hearer may feel different when he hear (1) and (2), in this case, typically on politeness of the speaker. The meaning of an utterance is not only the content of it, but a collection of all effects on the state of mind of the hearer. Sometimes, a speaker may select (1), the other time a speaker may select (2). This is determined from how polite he/she should be. When he/she want to be very poilte, the phrase *"of course you can change it"* **should not** be omitted.

So we conclude that what to omit is not determined only by the content it conveys, but also by the other factors such as politeness.

There is another reason why we think ellipsis is not always determined from the viewpoint of information which an utterance conveys. If an item which is identified from context should be omitted, it must be omitted in any languages. But there are differnces in ellipsis across languages. For example, ellipsis is used more often in Japanese than in English. This suggests there are some other factors which control ellipsis. We think politeness in one of the main such factors in English.

3 STRATEGIES TO GENERATE ELLIPTICAL ENGLISH UTTERANCES

We have collected 122 heuristics to generate appropriate elliptical utterance[5]. These heuristics are based on an analysis of dialogues from a text of English conversation for travellers[6]. They include dialogues between a Japanese traveler and a stewardess, an immigration officer or a hotel receptionist. Because relationship between them is so clear, politeness is kept clearly in their dialogues.

We investigate relations between elliptical expressions and politeness a hearer feels to the expressions. A native Enlish speaker helped the analysis.

From the result of our analysis, we classified politeness of utterances into four levels (level 0 to 3). Utterances at politeness level 0 are the simplest utterances that convey information the speaker intend correctly. These are the least polite utterances. Utterances at politeness level 3 are the most polite utterance. In above example, utterance (1) is the most polite (level 3) and (2) is the most impolite (level 0). Of course, there exist possible utterances at medium levels between them as follows.

(3) *Yes, of course you can.*

(4) *Yes, of course.*

In general, the more items omitted, the less polite an utterance becomes. But leaving too many items sometimes makes the utterance redundant. For example, consider the following two utterances as replies to a question *"May I have a blanket and a pillow?"*.

(5) *Certainly.*

(6) *Certainly you may.*

(6) seems to be a little redundant.

To generate appropriate elliptical utterances at an appropriate politeness level, Our heuristics determine how politeness level changes if some specific constituent is omitted.

Some of heuristics depend on some specific expressions such as *"would you"* or *"here/there"*. The other depend on syntactic structures such as "verb + pronoun" or "relative clause".

We show some of heuristics with examples. The utterances at some differnt levels can be the same as in Heuristics on "Yes" below because 4 levels are set tentative.

Heuristics on "Thank you / Thanks" "very much", "so much" and "a lot" in "Thank you ..." can be omitted at politeness level 2.

level 3	*Thanks a lot, here you are.*
level 2	*Thanks, here you are.*

Heuristics on "Yes" "Yes" cannot be omitted in short utterances (no longer than four words). In longer utterances it can be omitted at politeness level 1.

level 2	*Yes, he does.*
level 1	*Yes, he does.*

level 2	*Yes, I'd like to have it.*
level 1	*I'd like to have it.*

But if *"yes"* precedes a more complex clause such as a compound sentence, it can be omitted even at politeness level2.

level 3	*Yes, it's a large but very quiet room.*
level 2	*It's a large but very quiet room.*

There are some other heuristics including divergence like this.

Heuristics on relative clause Relative clause can be omitted at politeness level 3. But if the relative pronoun is *"that"*, the phrase after *"that"* can be omitted at level 3, but *"that"* can be omitted at level 2 or lower.

level 3	*I didn't know that (I couldn't do it).*
level 2	*I didn't know.*

Heuristics on "last noun phrase" Heuristics are applied recursively to an utterance until there are no heuristics applicable left. If only a noun phrase is left as a result of application of heuristics, the article of the noun phrase can be omitted at politeness level 1.

level 3	*May I have a blanket and a pillow?*
level 2	*A blanket and a pillow?*
level 1	*Blanket and pillow?*

4 IMPLEMENTATION

We have implemented a prototype of a system which generates elliptical English utterances based on 122 heuristics shown in the last chapter.

Utterances without ellipsis and politeness level are input to the system. The system determines which phrase can be omitted at the given politeness level and outputs elliptical version of the utterance.

First the input utterance is parsed with Context-Free Phrase Structure Grammar (CF-PSG) because some of our heuristics need syntactic information to plan ellipsis. Then which part of the utterance can be omitted at the given politeness level is checked with heuristics. Heuristics are applied recursively to an utterance until there are no heuristics applicable left.

Consider the following dialogue. (T: Japanese Traveler, S: Stewardess)

T: *Where is my seat?*

S: *It's over there facing the window.*

T: *Thank you very much.*

S: *You're welcome.*

In this example, *"It's"* and *"very much"* can be omitted at level 1 and 2 respectively. Then, if level 3 is given for the stewardess and 2 for the traveler, Output of the system is as follows.

T: *Where is my seat?*

S: *It's over there facing the window.*

T: *Thank you.*

S: *You're welcome.*

If level 1 is given for the stewardess and 3 for the traveler (odd case), Output of the system is as follows.

T: *Where is my seat?*

S: *Over there facing the window.*

T: *Thank you very much.*

S: *Welcome.*

5 EVALUATION OF THE SYSTEM

5.1 An Evaluation Test

To investigate the generality of our heuristics and system, we conducted an evaluation test to the system. Two sets of inputs are used, one is training data which

Table 1: Result of Evaluation Test (TR: training data, TE: test data)

	TR	TE
appropriate	92%	83%
inappropriate	8%	10%

is used for analysis. The other is test data containing dialogues among Japanese traveller, a doctor, a foreigner and a receptionist. We check outputs of the system if they are correct English utterances.

Table 2 shows the result of the evaluation test. With test data, the system generates appropriate ellipsis at over 80% ratio.

5.2 Failure Cases

There are two types of failures.

One is "over-ellipsis", that is, to omit an item that cannot be omitted as the following example.

S: *How feel now?*

Correct utterance at politeness level 1 is *"How do you feel now"*, but the system omit *"do you"* according to "Heuristics on "Do you".

The other type of failure is "to-be-omitted", that is, not to omit an item that should be omitted as the following example. (F: foreigner)

T: *Would you call him for me?*

F: *No, I wouldn't so that you have to go to the clinic.*

In F's utterance *"i wouldn't"* should be omitted at politeness level 1. But the system does not omit it.

The reasons of these failures are that our heuristics does not always cover all expressions and that some conflicts between heuristics may exist.

6 CONCLUSION

We proposed some strategies to generate appropriate elliptical English utterances, given some politeness level. Then, we showed outline of a system which generate elliptical utterances. The system can translate utterances with no ellipsis into appropriate elliptical utterances using the strategies. To use the system as a final module of natural language generation system, one can get appropriate elliptical utterances. From the result of the evaluation test with test data, the system generates appropriate elliptical utterances at over 80% ratio.

Our heuristics are based on only a few language data. To sophisticate them, we are going to analyze more data. We also plan to develop techniques to generate elliptical utterances in other languages. There seems to be differences in factors controlling ellipsis across languages. For example, politness is expressed by honoric expressions rather by ellipsis in Japanese. We must take some other factors into consideration to generate Japanese elliptical utterances.

ACKNOWLEDGEMENT

We would like to thank Mr.Roskoski at Faculty of Language and Culture, Nagoya University, who kindly helped us to analyze dialogues.

REFERENCES

[1] Clark H.H. and Schaefer E.F.: Contributing to Discourse, *Cognitive Science,* **13**, pp.259-294 (1989).

[2] Sanford D.L. and Roach J.W.: Representing and Using Metacommunication to Control Speaker's Relationships in Natural-Language Dialogue, *International Journal of Man-Machine Studies,* **26**, pp.301-319 (1987).

[3] Kohji Dohsaka: Interpretation and Generation through Efficient Representation, *IPSJ SIG-NL notes,* **82** (1991).

[4] Susumu Akamine and Osamu Furuse: A Japanese Generation based on Word Ordering and Ellipsis, *in Proceedings of the 49th Annual Convention IPS Japan,*4K-1 (1994).

[5] Wataru Sugimoto, Yuji Sagawa, Noboru Ohnishi, Noboru Sugie: Generating Elliptical Utterances — From a Viewpoint of Politeness —, *in Proceedings of Natural Language Processing Pacific Rim Symposium '95,* pp.414-419 (1995).

[6] JTB (Nihon Koutsuu Kousya): Hyaku-man nin no kaigai ryokou eikaiwa, JTB no kaiwa-syuu (1994).