A Framework for Modeling and Evaluation of Spoken Dialogue

Hiroya FUJISAKI, Sumio OHNO
Department of Applied Electronics, Science University of Tokyo
2641 Noda, 278 Japan
e-mail: fujisaki@te.noda.sut.ac.jp

1. Introduction Spoken language is clearly the primary means of human communication, and will no doubt be the major means of man-machine communication in the future. Compared with the written language, however, messages of the spoken language are less well-formed, with frequent occurrence of omissions, repetitions, word-order changes, errors, etc. Yet these so-called ill-formed features do not seem to present any obstacles to a human-to-human spoken dialogue. This leads us to believe that these features possess their own reason of existence and are actually utilized by the participants of a spoken dialogue to facilitate smooth and efficient information exchange. From this point of view, the present study analyzes human spoken dialogues and then tries to construct a model for the process of a spoken dialogue, with an aim of its utilization in human-machine dialogues.

2. Spoken Dialogue as a Purposeful Exchange of Information

Spoken dialogue is a purposeful exchange of information through spoken language among agreed participants. In many cases, the flow of information is globally unidirectional, but is often locally bidirectional. Participants generally possess a certain amount of knowledge and expectation on the knowledge, ability, and intentions of the other participants.

2.1 Prerequisites to a spoken dialogue

In order that a spoken dialogue should be possible, participants must observe the following general principles both in format and content of exchange.

(1) Use of relevant format — The exchange should follow certain rules of relevance in formant. For example, greetings should be returned, questions have to be answered, etc., though there are exceptions.
(2) Use of relevant contents — The contents of an utterance should be correct and relevant. (This does not hold in real life if it is against the interest of participants, but we will not deal with such situations here.)

2.2 Basic constraints of a spoken dialogue

(1) Parsimony — For the purpose of prompt communication, information that is judged to be redundant in view of the context, situation, listener's knowledge as well as the ability of reasoning, is completely omitted or extremely simplified, hence frequent use of ellipses, anaphora, etc.
(2) Real-time production — Prompt, real-time exchange of information does not allow the speaker to construct the whole message in advance, hence frequent occurrence of hesitations, filler sounds, re-starts, etc. These irregularities, however, do reflect the process of information processing on the part of the speaker, and may help correct decoding of the received message on the part of the listener.

2.3 Principles and methods for a smooth spoken dialogue

(1) Principles of cooperation — In order to ensure smooth and prompt exchange of information, participants use the following methods:

a. The speaker tries to select the information that is most relevant to the listener.
b. The speaker tries to select the expression that can be most easily understood by the listener.
c. When the listener's utterance allows multiple interpretations, the listener tries to select the one that is most likely to be intended by the speaker.
d. The listener tries to infer the speaker's intention that may not be explicitly stated.

(2) Principle of reliability — In order to ensure the overall reliability of communication, the speaker uses specific methods for the following purposes:

a. Drawing attention using prominence and repetition.
b. Confirmation of the listener's comprehension.
c. Indication of the speaker's own information processing stages.
d. Indication of retention, transfer, and request of speaking turns.
e. Repair of errors both on the part of the speaker and of the listener.

3. Linguistics of Spoken Dialogue Although speech used in the spoken dialogue is largely spontaneous and its variations are certainly greater than those of read speech, it has also its consistency as a linguistic system with its own units, rules and constraints, and any system of spoken dialogue processing has to take advantage of them. For instance, certain abbreviations, not found in conventional dictionaries, are quite consistently used in conversational speech so that they should be treated as lexical items. Likewise, certain sentence types, though they may look ill-formed from the point of view of grammar of written language, are quite consistently and reliably used in conversational speech. Furthermore, in purposeful spoken dialogues, the utterances are pragmatically more constrained than those of monologues and sentences in texts, since they have to be consistent with the goal of a particular dialogue. This regularity, which is largely unnoticed, is what makes the restoration of ellipses possible. A pilot study of the regularity and consistency is being conducted and will be summarized.

4. Analysis and Restoration of Ellipses in Spoken Dialogue

In order to investigate the possibility of automatic restoration of ellipses, elliptic expressions were classified into two categories: (A) first- and second-person pronouns (i.e., "I" and "you"), which are almost always omitted in dialogues, and (B) all others for which the corresponding words can usually be found in the preceding context.

4.1 Restoration of words of Category (A)

Four types of information available in the same sentence have been found to be useful in restoring ellipses of this category. They are:

(1) Inference based on polite expressions
Polite expressions belong to either one of the following two categories:

a. Honorific expressions — always refers to the second person.
b. Humble expressions — always refers to the first person.

(2) Inference based on verb attributes
Certain verbs such as 'think,' 'understand,' etc. refer only to the first person when they are used in the affirmative, while they refer only to the second person if they are used in the interrogative.
(3) Inference based on sentence types
Sentence types indicating command, request, exhortation, etc. are always used for the second person.

Of 99 ellipses found in 71 sentences from the text of simulated dialogues, 65% can be restored by these methods. Restoration of the remaining 35% requires the use of contextual information outside the same sentence.

4.2 Restoration of words of Category (B)

There are three cases where the preceding context provide information for restoring ellipses. They are:

(1) Restoration using the word found in the preceding context
(2) Restoration using only a part of a compound word
(3) Restoration using only a part of a phrase of the type "A of B"

These cases constitute 94% of the 330 examples investigated, and the total success rate of restoration is 67% using only the immediately preceding utterance by the partner of the dialogue.

5. Evaluation of Goodness of Questions and Answers Investigation of the type of information exchange found in simulated dialogues revealed that a predominant part of purposeful spoken dialogues consists of questions and answers. It is thus quite important for a smooth and efficient spoken dialogue to ask 'good' questions and to give 'good' answers. It is quite common, that the ultimately requested information can be obtained only by repeated questions and answers. Let A be the person who seeks information and B be the one who supplies it. In actual situations, the roles of A and B are not fixed, but here we assume that they are fixed. Individual questions made by A may contribute only a small portion of the information which A seeks, but the series of questions posed by A will contain the information as to what A eventually wants, and will thus provide B with the information on A's intention.

Assuming two types (Yes/No-type and W·H-type) for A's question and three types (Simple, Comprehensive, and Optimum) for B's answer, we have six situations shown in Table 1.

The goodness of a question or an answer depends of course on the criterion as well as the situation. In a simple yes/no situation, the optimum question by A, from an information-theoretical point of view, is the one that will maximize the expected amount of information to be gained by the answer, and the problem of selecting the optimum series of questions is solved by the method of Huffman coding. This can be extended to the case of comprehensive answers. In both cases, it is assumed that B answers the question automatically. When B possesses the ability to infer A's intention, however, the optimum question by A will be the one that will facilitate B's inference on A's intention.

On the other hand, the goodness of answer by B in a simple yes/no situation is the accuracy. In the case of a comprehensive answer, it should at least provide the order of

Table 1. Classification of Questions and Answers.
(A: one who questions, B: one who answers)

Type of Questions by A	Type of Answers by B		
	Simplest	Comprehensive	Optimum
Yes / No	Yes or no or don't know	Provide all the conditions for yes and no	Add further information to meet A's intention
W · H	Give one answer for the default condition	Give all the answers with respective conditions	Indicate alternatives to meet A's intention

preference by some reasonable criteria. The optimum answer will be the one that will best fit A's intention if there is one, and will show the next alternative even if the best is not available.

Keywords: spoken dialogue, ill-formedness, parsimony, real-time production, principle of cooperation, principle of reliability, restoration of ellipses, goodness of questions