A cognitive science approach to the problem of dialogue: Roles of language information, speech information, and visual information in understanding of the dialogue partner

Jun-ichi Abe
Department of Psychology
Faculty of Letters
Hokkaido University
Sapporo 060, JAPAN
E-mail: abe@hubs.hokudai.ac.jp

In order to participate cooperatively in dialogue, we need to consider the dialogue partner and his/her situation. Necessarily, we need to obtain wide information about the partner. Suppose a face-to-face dialogue situation. Using the auditory modality, we can obtain the partner's speech sound information, and from that information we can extract the partner's language information. Using the visual modality, we can obtain the partner's facial and bodily information. We use such kinds of information to make a proper internal model of the partner and then to understand him/her.

I believe the following questions should be asked for clarifying the cognitive process of understanding the partner in a dialogue.

1) What part of the auditory and visual information is taken as cues for understanding the dialogue partner?

2) Using these cues, what kind of inference is made and how much detailed is the inference?

The present paper reports a part of our experimental research program which attempts to answer the above mentioned questions.

1. METHOD

1.1 Stimulus materials and design of the experiment

From TV and movies we recorded many scenes in which a cast is speaking toward the camera with his or her upper half of the body being pictured. From these recorded scenes, we chose 14 scenes (of 5 to 7 sec long) in which a speech is continuous and without a background music. These scenes were loaded on a computer to edit the contents of the scenes as follows.

Condition 1 (picture + voice): Original scenes.
Condition 2 (picture + caption): Delete the sound and superimposed the caption.
Condition 3 (picture only): Delete the sound.
Condition 4 (voice only): Delete the picture.
Condition 5 (caption only): Delete both the picture and sound and add the caption.
Condition 6 (picture + voice + caption): Add the caption to the original scenes.

Condition 7 (voice + caption): Delete the picture and add the caption.

1.2 Procedure

The subjects were asked to write down on a response sheet what they judge and infer about the speaker. The question items used are as follows: 1) What do you think the speaker feels when he is speaking. How could you judge the speaker's feeling? 2) What is your judgment on the speaker's character? How could you judge the speaker's character? 3) What would you guess the context in which this scene took place? How were you able to guess so?

The subject was required to make his or her judgment on the above three questions at each presentation trial and was required to record each of the judgments on the response sheet. At each occasion, the experimenter recorded the time taken for the subject to complete the response.

1.3 Subject

We recruited 28 undergraduate students at Hokkaido University. Four subjects were allocated randomly to each of the 7 types of the presentation conditions.

2. RESULTS AND DISCUSSION

The data were standardized and were submitted to the 2-way analysis of variance (7 presentation conditions x 14 scenes). With regard to the analysis of the number of characters written, the presentation condition yielded a significant main effect. The main effect of scene was also significant. However, the interaction was not significant.

Results of the multiple comparison showed significant differences between the following presentation conditions: condition 3 (picture only) versus condition 6 (picture + voice + caption), condition 4 (voice only) versus condition 1 (picture + voice), that versus condition 6 (picture + voice + caption), that versus condition 7 (voice + caption), condition 5 (caption only) versus condition 1 (picture + voice), that versus condition 6 (picture + voice + caption), and that versus condition 7 (voice + caption).

We expected the number of characters in response has a monotonic increasing relation to the amount of information received. The results confirmed this expectation: The number of characters responded was greater in conditions 1 (picture + voice) and 6 (picture + voice + caption), but fewer in conditions 3 (picture only), 4 (voice only) and 5 (caption only).

It is noteworthy that there was a significant difference between condition 4 (voice only)and condition 7 (voice + caption). This result was not expected. We speculated that adding captions to voice would not increase the amount of information per se. However, adding captions did contribute in increasing the number of characters written on a response sheet. We believe that this effect was caused by having the subject read the caption to facilitate his/her conceptual and schematic knowledge activation. The subject, thus, could have accomplished various inferences in a short amount of time.

As a result of analysis of response times, we found significant main effects of the presentation conditions and of the scenes. The interaction, however, was not significant. Result of the multiple comparison showed significant differences between the following conditions: condition 4 versus condition 6, condition 5 versus condition 1, that versus condition 2, and that versus condition 6. The difference of the response times between presentation conditions showed almost the same fashion as that of the numbers of characters on a response sheet. That is, the condition with greater amount of information tend to give rise to longer response times and vice versa.

The results, as a whole, suggest the uniqueness of written language as a source of information. We predicted that there is no difference between condition 6 (picture + voice + caption) and condition 1 (picture + voice) as far as the actual amount of information goes. However, both the response time and the response amount were greater in condition 6 than in condition 1. The results suggest that the visual language information in the form of characters has a potential to strongly activate the lexical, the conceptual, and the schematic knowledge.

Keywords: cognitive processes, dialogue, understanding, psychological experiment