Analysis of the Transitions of Spatial-Temporal Scenes in Everyday Conversation

Yoshiko Kawabata
NINJAL, Japan
kawabata@ninjal.ac.jp

Mikio Nakano
C4A Research Institute, Inc., Japan
mikio.nakano@c4a.jp

Abstract

In conversations, the participants need to imagine the place and time where events mentioned by other participants occur to understand utterances. In this study, we refer to this place and time information as *scene*. We have been analyzing the Corpus of Everyday Japanese Conversation (CEJC) to investigate how scenes are expressed in conversations, including visual information. This paper describes the concept of scenes and reports the clues to scene transitions found in our analysis.

1 Introduction

In conversations, speakers may refer to objects in front of them, events that occurred in a specific place in the past, or future plans. For successful communication, it may be necessary to understand the information about the time and place where events occurred. Example (1) is part of a conversation included in the Corpus of Everyday Japanese Conversations (CEJC) (Koiso et al., 2022). [1] In (1), friends Naoya and Yumiko are talking. Yumiko is on a restricted diet due to illness and is talking about a shopping trip to the supermarket with her sister. To comprehend the expression "putting ice cream" in line 5, which refers to an event in a supermarket, it is necessary to understand the place where the event occurred. This study refers to the information about time and place necessary to understand such a conversation as **scene in a conversation** (hereafter abbreviated as *scene*). This paper presents our ongoing research in which we have been analyzing the CEJC to investigate how scenes are expressed and shift in conversations.

---

[1] Originally in Japanese. "(L)" denotes laughter, and numbers inside "{ }" indicate pause seconds. Transcription symbols irrelevant to the discussion in this paper have been omitted.

(1)

1	Yumiko	So, **yesterday** or the day before, I went to the **OK Store** with my sister.	
2	Naoya	Yeah, yeah, yeah.	
3	Yumiko	And since my sister is healthy, she	
4	Naoya	(L)	
5	Yumiko	was putting ice cream and stuff, saying, "Don't look."	
6	Naoya	Yeah. (L) Yes.	

(CEJC:T002_015 1623.571-1633.426)

2 Scenes in Conversation

In (1), the participants are considered to be cognitively processing the scene where Yumiko and her sister were at the supermarket one or two days ago and imagine Yumiko's sister is putting ice cream into the shopping basket. This kind of processing is part of the conversation understanding process and is necessary for comprehending subsequent utterances. We regard scenes as part of the mental model held by conversation participants (Bower and Morrow, 1990). Scenes have at least two elements: time and place. Time can be further distinguished into categories such as present, past, future, and hypothetical, while place can be distinguished between what is in front of the participants and other locations. In conversations, either or both of these elements can be unspecified, but for successful communication, the participants must share the scene to an extent sufficient for the conversations.

3 Transitions of Scenes in Conversation

In conversations, scenes frequently shift, so participants need to recognize these transitions. This section reports on the cues used for scene transitions. One cue is the use of explicit time (e.g., yesterday, the day before) and place expressions (e.g., OK Store), as seen in the first line of (1).

The second cue is changes in the surrounding physical environment. Notably, (2) is a conversation between a married couple while watching TV (Figure 1). At the time of the first and second lines, different people appear on the TV screen, and the

couple refers to the individuals shown. Their communication is successful because they are observing the same external environment and sharing the scene's changes. Such immediate scene transitions also occur during conversations while performing tasks or traveling in a car.

Figure 1: The husband (on the left) and wife (on the right) are watching TV. The TV is showing the news, and a member of the parliament is on the screen.

(2) <A young politician (Otokita) appears on TV>
  1 W Otokita-kun is doing his best.{2.39} Good for him.{35.82}
     <The screen changes to show an older politician>
  2 W Hmm. I feel like he's about to die.
  3 H Somehow, the ending isn't very good.
  4 H This person.
  5 W Hmm.
                     (CEJC:C002_003 213.092-262.987)

The third cue is the use of participants' body movements. Example (3) is part of a casual conversation among five female friends. In the first half (lines 1-6), they are talking about the weather in Tokyo just before the conversation. However, Kanako's remark in line 8, "You want to say it's dark," does not refer to the weather but rather to her skin color. Understanding this remark through language alone is difficult, but it becomes clear when observing the body movements. Just before Misaki calls Kanako's name in line 7, she moves her arm close to Kanako's arm and looks at them. Kanako notices this action by Misaki (Figure 2). By using such body movements, the participants' gaze is directed to the present object, shifting the scene from past weather (not in front) to the immediate present.

(3) 1 Natuko But, you know, it cleared up by noon, right?
     2 Misa   It cleared up.
     3 Reiko   Yeah yeah, yeah yeah.
     4 Misa   Yeah. Right.{1.29}
     5 Mika   Really?
     6 Reiko   I see.
        <Misa looks at Kanako's arm, and Kanako notices it (Figure 2)>
     7 Misa   Kana-chan.
     8 Kanako You want to say it's dark.
     9 Misa   We're both wearing short sleeves.
    10 Kanako Yeah.
                     (CEJC:C001_001 660.06-673.868)

Figure 2: Misa (on the right) moves her arm closer to Kanako (on the left), and they are both looking at each other's arms.

4 Related Work

Temporal and spacial information is considered one of the factors that define segments in discourse structure (Hobbs, 1978; Asher et al., 2007; Hoek et al., 2019). While the relationships between segments are basically between a given segment and the preceding text, Charolles et al. (2005) propose the notion of *discourse frame* as a structure that affects the interpretation of the subsequent text. According to the concept of discourse frames, Yumiko's utterance in (1) can also be seen as a frame realized by the temporal and spatial expressions in the first line. The concept of the scenes in this study is an interpretative framework constructed by including information about the surrounding environment and body movements, and it can be seen as an extension of the concept of discourse frame.

5 Future Work

We plan to clarify the relationship between scenes and linguistic expressions. However, while analyzing the CEJC, we often encounter difficulties in determining scenes. Accordingly, we aim to examine the criteria for making such judgments, including the definition of the scene itself. Additionally, we intend to clarify the distinction between scenes and other conversational-structure-related concepts such as purpose and topic.

Acknowledgments

References

Nicholas Asher, Laurent Prévot, and Laure Vieu. 2007. Setting the background in discourse. *Discours*, 1.

Gordon H. Bower and Daniel G. Morrow. 1990. Mental Models in Narrative Comprehension. *Science*, 247(4938):44–48.

Michel Charolles, Anne Le Draoulec, Marie-Paule Pery-Woodley, and Laure Sarda. 2005. Temporal and spatial dimensions of discourse organisation. *Journal of French Language Studies*, 15(2):115–130.

Jerry R. Hobbs. 1978. Why is discourse coherent? Technical Report 176, SRI International, Menlo Park, CA.

Jet Hoek, Jacqueline Evers-Vermeul, and Ted J. M. Sanders. 2019. Using the Cognitive Approach to Coherence Relations for Discourse Annotation. *Dialogue & Discourse*, 10(2):1–33.

Hanae Koiso, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yasuyuki Usuda, and Yuka Watanabe. 2022. Design and evaluation of the corpus of everyday japanese conversation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5587–5594.