神戸市外国語大学 学術情報リポジトリ

Taking turns at talk on screen : an analysis of the correlation between words and images in cinematic discourse

メタデータ	言語: eng 出版者: 公開日: 2013-03-06 キーワード (Ja): キーワード (En): 作成者: Kizu, Kumiko メールアドレス: 所属:
URL	https://kobe-cufs.repo.nii.ac.jp/records/1475

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

Taking turns at talk on screen:
An analysis of the correlation between words and images in cinematic discourse

Kumiko KIZU

This study investigates how language and images are combined in cinematic discourse. The analytic focus is on two-party conversational scenes presented through the shot/ reverse-shot exchange by which two characters are shown in close-up on screen one after the other. The analytic methodology involves a transcription model for the combinations of audio and visual resources from released films (DVDs), and the concepts of conversation analysis (CA) (Sacks, Schegloff & Jefferson, 1974 etc.). As a result of the investigation, this study proposes an essential correlation between words and images in cinematic discourse, and it argues that the combination of character utterances and their images is motivated by the communicative need to facilitate the audience's comprehension of the interactive negotiation between the characters on the screen where the audience cannot participate. Furthermore, this study addresses the relationship between the communicative function that is achieved with the combination of the two media and the film-specific structure.

Cinematic discourse is multimodal: films are composed of multiple representational and communication resources including language, images, music and sound effects. In addition, films are created through three phases (Monaco, 2000): pre-production (scriptwriting, casting, etc.), shooting, and post-production (editing and sound mixing). The films we usually watch in the movie theater or at home are finalized by combining the multiple elements which have been created separately in time and place. Furthermore, cinematic discourse includes two communicative levels (Yamaguchi, 1998): the micro-cosmic communication between characters in the fictional world and the macro-cosmic communication from the filmmaker to the audience in the real world. Given such a complex construction, different modes are carefully organized in a way that the audience is able to grasp the story and enjoy the fictional world that a film presents (Ondaatje, 2002). Thus, it is crucial to discuss the multimodal aspects and the correlation among different media, with the audience taken into account, in order to fully understand cinematic discourse.

The traditional studies in cinematic discourse, however, have not sufficiently addressed the matter of multimodality. They have focused either on the visual/spatial aspect or on the audio/verbal aspect, referring to the other aspect in passing when needed. Recently, an interest in cinematic multimodality has surged. Unfortunately, however, few systematic explanations

have been provided about the essential ‘interrelation’ among modes, although there has been substantial analysis of how the ‘integration’ of different modes affects the audience’s comprehension of scenes. In addition, the relation between the audience’s interpretive process and the fundamental structure of film has not been carefully explained. Cinematic discourse should be discussed theoretically as well as empirically, based on both the human’s cognitive process of modes and the film-specific way of combining modes.

In an attempt to take the ideal approach to cinematic multimodality mentioned above, this study concentrates on the two media of spoken language and images, and specifically focuses upon a simple combination of these two media: the combination of two characters’ utterances and their images in conversational scenes presented by the shot/ reverse-shot exchange. In such conversational scenes, the moment one character appears on the screen is not always synchronized with the moment s/he starts to talk. Thus, there is, if not always, a time lag between the shot transition and turn transition. In this time lag, one character’s voice is combined with the other’s face (i.e., the silent character’s close-up). In other words, the other remains on screen as the hearer if the shot transition occurs a little or long after turn transition, or s/he appears on screen as the hearer if the shot transition occurs slightly or long before turn transition. This study demonstrates that the timing of the shot transition is designed to show the audience the way characters take their turns—smoothly, early or late, and it argues that the combinations of character utterances and their images provide the frames for the audience to interpret and/or make inferences from the interactive negotiation between the characters on the screen. CA concepts are borrowed to explain the way the timing of the shot transition is manipulated to show the dynamics of the interaction between on-screen characters. The findings of everyday interaction are employed to demonstrate how fictional conversation is designed based on the practices and knowledge in everyday conversation, which the filmmaker assumes to share with the audience.

From the viewpoint of the structure of cinematic discourse, the design of conversational scenes presented by the shot/ reverse-shot exchange also possesses great significance: it not only constitutes an elementary combination pattern of language and images in cinematic discourse (the combination of one verbal variable and one visual variable), but also represents a basic form of the film structure. As the previous film theorists point out (Monaco, 2000 etc.), film is characterized and differentiated from other representational arts, such as novel and drama, by the two major film-specific techniques—camerawork and editing. Scenes are constructed to lead the audience to the intended interpretations by limiting what the audience sees in a shot (camerawork) and by controlling when they see or stop seeing it and in what order (editing). In conversational scenes presented by the shot/ reverse-shot exchanges, a close-up shot extremely

limits what the audience sees to one character's face, and the shot alternation exclusively controls when the audience sees and stops seeing the character's face. Thus, such conversational scenes also constitute a very fundamental form of the structure of cinematic discourse. Therefore, these conversational scenes are suitable for investigating the essential correlation between words and images in cinematic discourse.

Based on these analytic viewpoints and background, this paper proceeds as follows:

In Chapter 2, the previous studies in cinematic discourse are introduced, and the viewpoint and significance of this study are clarified. First, I summarize monomodal approaches to cinematic discourse which focus on either visual or verbal aspects. Second, I refer to two groups of studies which are immediately related to the current study: the first one is about multimodality; the second one is about the inferential frame which is designed to help the audience to interpret and/ or make inferences from scenes.

In Chapter 3, data and methodology are introduced. First, I introduce the twelve dual-protagonist films from which I have collected data of the combination of character utterances and their images. Then, I describe the shot/ reverse-shot exchange. Subsequently, I summarize CA concepts employed in this study: turn-taking, adjacency pairs and preference, and I examine the work on conversational rhythm and gaze direction which are the main variables in the analysis of audio and visual modes.

Analysis starts in Chapter 4. First, I report on the four types of combinations of two characters' voices (utterances) and faces (close-ups) that are found in my data: the synchronized type (the start of one character's voice synchronizes with his/her face's appearance on screen, therefore no time lag occurs); the voice-first type (one character's voice precedes his/her face on the screen); the face-first type (one character's face precedes his/her voice on the screen); and the mute type (one character does not take his/her turn on the screen). Then, I further divide the voice-first type and the face-first type into three subtypes according to the duration of the time lag: the voice-first type 1 and the face-first type 1 (the basic combination types that have a short time lag); the voice-first type 2 and the face-first type 2 (the shorter complex combination types that have a relatively short time lag, though longer than the basic combination types); the voice-first type 3 and the face-first type 3 (the longer complex combination types that have a relatively long time lag), and I examine the linguistic features of the words that are combined with the hearer's image in each combination type.

In Chapter 5, the focus of the analysis is on the way the time lag occurs. I analyze the three basic combination types (the synchronized type, the voice-first type 1 and the face-first type 1), and I explore the primary function of the combinations of words and images based on the findings of interactional rhythm in everyday conversation (Couper-Kuhlen, 1993 etc.). As a

result of the analysis, I argue that the combinations of the two media primarily function as contextualization cues, indexing the timing in which the next speaker starts his/her turns—integrated, anticipated or delayed. I also argue that the combinations of words and images provide the interpretive frames that help the audience to infer the next speaker's attitudes and the relationship between the characters on screen. Furthermore, I argue that these functions are film-specific in that they lead the audience to an intended interpretation, independently of the "actual" way the characters take their turns in the fictional world.

In Chapter 6, the analytic focus is shifted to the duration of the time lag. I analyze the shorter complex combination types (the voice-first type 2 and the face-first type 2) and the longer complex combination types (the voice-first type 3 and the face-first type 3), and I explore the durational function of the combinations of words and images. Here, the hearer's gaze direction is scrutinized based on the findings in everyday conversation (C. Goodwin, 1979 etc.). As a result of the analysis, I conclude that the duration of the time lag iconically indicates the complexity of the interaction between the characters on the screen: the more complex the interaction between the two characters becomes, the longer the time lag becomes, so that the audience easily and clearly perceives how the hearer acts or reacts to the speaker's utterance. The hearer's image is necessary for the audience to understand how the two characters negotiate with each other over a turn on the screen.

In Chapter 7, I move on to analyze the remaining combination type, or the mute type. In this type, the length of the time lag is maximized since no turn transition occurs at and around the shot transition. Here, the analytic focus is on the arrangement of the words that are combined with the hearer's image in this long time lag. I point out that the words may be arranged in order from retrospective ones to prospective ones in a shot. I argue that such an arrangement of words signals to the audience a point at which the hearer should respond to the utterance, and that it also functions to facilitate the audience's comprehension of the interaction between the characters on screen.

Finally, Chapter 8 discusses the correlation between words and images in cinematic discourse based on the findings of the analyses in the previous chapters. First, I give a systematic explanation of all the combination types that occur in conversational scenes presented by the shot/ reverse-shot exchange. Then, I propose an essential correlation between language and images in cinematic discourse as follows:

- The speaker's image tends to be combined with informative words in a turn;
- On the other hand, the hearer's image tends to be combined with less informative but interactively important words in a turn;
- These interactively important words include two types—retrospective and prospective ones:

the former is combined with the image of the hearer who does not start his/her turn; the latter is combined with the image of the hearer who has finished his/her turn;

- When the character on screen does not take his/her turn, these interactively important words are arranged in order from retrospective types to prospective ones.

Thus, the combination of words and images is motivated by the communicative need to facilitate the audience's comprehension of conversational scenes, and it provides the frame that the audience relies on to interpret and/or make inferences from the interaction between the characters on screen: the combination of retrospective words and the hearer's image signals that what the speaker has said before the shot transition should be interpreted based on how the hearer reacts, and then, the combination of prospective words and the hearer's image signals that what the speaker is going to say after the shot transition should be interpreted based on what the hearer expects.

Subsequently, the relationship of the inferential frames that are provided by the combinations of character utterances and their images, and the film-specific structure is discussed. I argue that such inferential frames are designed to create a gap, incoherence or deviation that invites the audience to make inferences from the interaction between the characters on screen, based on the basic procedures of the film editing (Asanuma, 1990).

This study is by no means exhaustive, and it narrows its analytic target and variables in order to extract an essential correlation between language and images in cinematic discourse. This study, however, demonstrates an approach to tackle cinematic discourse which has diverse aspects to be analyzed. The analytic perspective of this study has possibilities as a basis for investigating cinematic conversation in other languages, such as Japanese which has different interactional resources than English (Szatrowski, 1997, Hayashi, 2001, etc.). Furthermore, given that cinematic discourse is designed based on the practices and knowledge in everyday conversation, I believe that this study on fictional conversation will provide an insight into how we interact with each other in the real world.