Take the words out of my mouth: Verbal instructions as embodied practices

Anja Stukenbrock *

University of Duisburg-Essen, Fakultät für Geisteswissenschaften, Germanistik/Linguistik, Universitätsstr. 12, 45117 Essen, Germany

Abstract

This paper is concerned with key concepts in conversation analytic research: projection, conditional relevance and the adjacency pair. Focusing on a specific type of adjacency pair (instructing/instructed action) that constitutes an exemplary case for the analysis of embodied practices, the paper examines how verbal and embodied resources are mobilized, temporally organized and oriented to by the participants to form instructing and instructed actions.

A recurring practice in my data is the combined use of the German deictic form ''so'' with concurrent bodily demonstrations. In instructional sequences, the expression ''so'' not only serves as a turn-internal flag (Streeck, 2002) that directs the addressee's visual attention to the speaker's bodily activities, but it can also be part of a projecting first (instruction) or of a projected second pair part (instructed action). Drawing on video recordings from different settings, it will be shown how demonstrations carried out by the instructing person on the one hand and by the instructed person on the other hand are embedded within the sequential format of instructions. The paper argues for a multimodal conceptualization of central notions of CA such as turn, adjacency pair, and conditional relevance. It also raises the theoretical question of how the concept of conditional relevance as it defines the classical adjacency pair relates to the concept of projection when both are (re)considered from a multimodal perspective.
© 2013 Published by Elsevier B.V.

Keywords: Embodied interaction; Adjacency pair; Conditional relevance; Projection; Instruction; Modal deixis

1. Introduction: doing things with your body

As a fundamental quality of face-to-face interaction, multimodality is particularly important for instructional sequences: In co-present situations, instructing people (how) to do things, (how) to do things with things and (how) to do things with their body requires not only speech, but crucially involves the use of visual bodily resources.

In instructions that deal with bodily practices, the human body becomes relevant in a double way. Firstly, embodied practices are mobilized as resources in interaction, they form an integral part of utterance construction in face-to-face communication and play a fundamental role in establishing, maintaining, and dissolving units of interaction (Goodwin, 1980, 2003, 2007; Kendon, 1990, 2004; Mondada, 2009b; Schegloff, 1984, 1998; Streeck, 1993, 2002, 2009).

Secondly, in cases where the instruction concerns a bodily practice, i.e. some kind of activity which requires special motor skills, professional techniques, the handling of tools and objects, etc., bodily practices moreover constitute the object of communication, demonstration, and assessment. They become thematically relevant and are made available by participants for visual inspection of their co-participants. In order to be taught and learned, embodied practices (bodily objects of instruction) might need to be demonstrated by the instructing person and imitated or exercised by the instructed person.

---

* Tel.: +49 201 183 4168.
  E-mail address: anja.stukenbrock@uni-due.de.

In conversation analytic and ethnomethodological literature, it has been observed that instructions are sequentially organized as adjacency pairs. The first pair part, termed ''instruction'' (Garfinkel, 2002) or ''directive'' (Goodwin, 2006; Mondada, 2011), establishes the conditional relevance for a second pair part, termed ''response'' (Goodwin, 2006), ''instructed''/''following'' (Garfinkel, 2002) or ''complying action'' (Mondada, 2011). It has also been noted that other responses than verbal utterances can play a decisive role (Lindwall and Ekström, 2012; Mondada, 2009a, 2011). Examining instructional sequences in video-recorded driving lessons, De Stefani and Gazin (in this issue) observe that the complying action is often delivered bodily: ''While the instruction is always produced through talk, the projected action is systematically achieved multimodally''.

Rauniomaa and Keisanen (2012) discuss two multimodal formats of responding to requests. Whereas in the second format, the response is made up of a (verbal) acceptance and a subsequent fulfillment of the request, the response in the first format consists of a fulfillment only. In their data, the requests often deal with material objects or ongoing activities. The responding fulfillment is therefore carried out bodily. Similarly, in a single-case analysis of an instructional sequence involving manual skills, Lindwall and Ekström (2012) describe embodied actions made relevant by an ongoing reciprocal calibration of instructing and instructed actions. They argue that instructions-in-interaction strengthen the claim ''that the notion of 'adjacency' or 'nextness' should not be limited to turns-at-talk'' (Lindwall and Ekström, 2012:36).

Based on a set of video data from a range of settings (cf. section 2), I will show that multimodality is not only relevant for the instructed action, but also for the instructing or directive first pair part: In activity frameworks in which the object of instruction is a bodily practice, instructions are often delivered multimodally, i.e. by a bodily demonstration which invites the instructed person not only to listen, but to watch in order to discern relevant features of the embodied skill.

A recurring feature in my data are instructional sequences in which participants routinely use the deictic form ''so'' to project a visible bodily action and thus request the co-participant to pay online attention to an emerging body movement. As such, the expression ''so'' can be both part of a projecting first or of a projected second multimodal pair part. Studying the projective force of ''so'' in different sequential surroundings within instructional sequences, the aim of this paper is twofold: On the one hand, it studies empirically how verbal and embodied resources are temporally organized within and across turns to project and to fulfill instructions and instructed actions. In this context, it is also particularly interested in phenomena below the level of turns and actions which nevertheless have a kind of pair-like connection. On the other hand, it is theoretically concerned with the question of how the concept of conditional relevance as it defines the classical adjacency pair relates to the concept of projection when both are (re)considered from a multimodal perspective.

Following a brief note on data, transcription, and methodology in section 2, section 3 sketches some general problems concerned with the notion of multimodal adjacency pairs. It offers a sequential analysis of a seemingly simple example to show how projection and conditional relevance across modalities become relevant. In section 4, I will give a brief review of existing studies on the German form ''so'' and describe its function as a gesturally used deictic in multimodally delivered instructional sequences. The empirical analyses focus on bodily demonstrations performed by the instructing person to visualize the instruction (section 5), demonstrations done by the instructed person to ask for an assessment (section 6), and an extended sequence of demonstrations undertaken both by the instructing and by the instructed person to arrive at a shared embodied understanding of a particular practice (section 7). The paper concludes with a discussion of the empirical findings (section 8).

2. A note on data, transcription, and methodology

The analysis is based on various types of video data. The data comprise video recordings of everyday communication, self-defence trainings, doctor–patient interaction, a reality TV show and cooking shows.[1] The decision to use a wide range of different data is methodologically motivated: the choice of data overcomes the delimitations of a single setting and allows me to generalize my observations on a specific practice: the multimodal use of the deictic form ''so''. Furthermore, by covering informal as well as formal talk, everyday as well as institutional settings, dyadic as well as multi-party participation frameworks, stationary as well as mobile configurations, the data constitute a more reliable basis for a multimodal reconceptualization of the notions of *conditional relevance*, *adjacency pair* and *projection*. The collection comprises 20 cases of multimodally formatted instances of ''so'' (Stukenbrock, 2010). Not all of the above mentioned subcorpora will be represented in the examples discussed in this paper.

The transcription of talk is based on the GAT transcription conventions developed by Selting et al. (1998). The system has been revised and is now called GAT2 (Selting et al., 2009; for an English translation cf. Couper-Kuhlen and Barth-Weingarten, 2011). The conventions are explained in detail in Appendix. To represent the relevant bodily actions and

---

[1] The use of video material taken from TV broadcasts can pose a range of problems as this material is subject to the TV editors' practices of cutting, editing, and presenting the chosen material (cf. Broth, 2008). For the present discussion, the extracts have been carefully chosen with regard to the online visibility of the relevant phenomena.