

**Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule
of the turn-taking system of conversation analysis***

Peter Auer

Department of German/Section of German Linguistics & Hermann Paul Centre for Linguistics
(HPCL)

University of Freiburg, Germany

peter.auer@germanistik.uni-freiburg.de

postal address:

Prof. Dr. Peter Auer

Department of German

Albert-Ludwigs-Universität Freiburg

D-79085 Freiburg

Germany

**Accepted for publication in *Discourse Studies*.
Author's version, July 2020**

* I wish to thank Anja Stukenbrock and Geert Brône for most valuable comments on a previous version. My thanks also go to Katharina Dankwarth, Johanna Hantsch, Clarissa Weiß for their help with the coding of the examples.

Abstract: It is argued in this paper that a multimodal analysis of turn-taking, one of the core areas of conversation analytic research, is needed and has to integrate gaze as one of the most central resources for allocating turns, and that new technologies are available that can provide a solid and reliable empirical foundation for this analysis. On the basis of eye-tracking data of spontaneous conversations, it is shown that gaze is the most ubiquitous next-speaker-selection technique. It can function alone or enhance other techniques. I also discuss the interrelationship between the strength for sequential projection and the choice of next-speaker-selection techniques by a current speaker. The appropriate consideration of gaze leads to a revision of the turn-taking model in that it reduces the domain of self-selection and expands that of the current-speaker-selects-next sub-rule. It also has consequences for the analysis of “simultaneous starts”.

Bioblurb

Peter Auer is Full Professor of German Linguistics at Albert-Ludwigs-Universität, Freiburg (Germany). He has also taught German and General Linguistics at the universities of Konstanz, Saarbrücken and Hamburg. His research has been devoted to a wide range of fields in linguistics including bilingualism, social dialectology, phonology, syntax of spoken language, conversation analysis, patholinguistics and sociolinguistics. He has published more than 200 scholarly articles in journals and edited volumes, and is the author various books.

1. Introduction

Technological innovations can change our professional way of looking at human communication in general, and linguistic interaction in particular, by enabling us to gather new kinds of data and analyze them in novel ways. Entirely new fields with their methodologies have emerged from technical innovations. A good example is conversation analysis, if “playing around with tape-recorded conversations, for the single virtue that I could replay them” was indeed one of the reasons that led Harvey Sacks to study conversation (lecture September 3, 1967, p.7). The tape recorder had already existed for some decades by the mid-1960, but in the 1960s they became cheaper, relatively small (and therefore unobtrusive) and easy to handle. Sacks understood that beyond its technological advantages, the tape recorder made it possible to look at conversational interaction in a completely new way; he turned the tape recorder into a conversation-analytic tool, a kind of microscope for interaction analysts.

In this article, I want to demonstrate that the analysis of naturally occurring human interaction can today again profit from new technologies. They may not be quite as easy to handle as tape recorders were in the 1960s, but they have recently become sufficiently advanced and unobtrusive to be used outside experimental settings. The technology I have in mind here is mobile eye-tracking, and the focus is on the conversational turn-taking system as investigated by Sacks, Schegloff & Jefferson in their groundbreaking 1974 article. I argue that the question of how a current speaker can select a next speaker is so fundamentally linked to human gaze that any account of turn-taking that does not include it will run the risk of remaining incomplete and even be misleading.

2. Eye-tracking as a method for studying spontaneous face-to-face interaction

The basic idea of eye-tracking is to record foveal vision, i.e. the field of vision in which visual acuity is at its highest. Eye-tracking has become one of the standard techniques of experimental research, particularly in psycholinguistics and cognitive science. Nowadays, however, it can be used for interactional studies as well, as mobile, head-mounted trackers are available which, although somewhat less precise than stationary ones, allow the speaker to move freely and gaze naturally at their surroundings (see Brône & Oben, eds., 2018 for more details and applications). In addition to the infra-red camera that films the movements of the pupil, head-mounted eye-tracking glasses include high-resolution scene cameras. The scan path of the tracker can therefore be superimposed on the picture of the scene camera. As an illustrative example, Fig. (1) shows a split screen with four synchronized recordings: the external camera's view of the three people sitting around a table, each of them equipped with a eye tracking glasses, and each of these participants' scene camera with the areas of their focal vision marked by a 'cursor' (circle). As can be seen, the person sitting on the left side (whose view is displayed in the lower left part) and the person sitting in the middle (displayed in the upper left) are gazing at the third person (displayed in the upper right) at this particular moment. This gazed-at person looks at the woman in the middle, i.e. there is mutual gaze between the two.



Fig. (1): Split screen representation of eye-tracking recording with three participants

Eye-tracking of this kind enables the analyst to work with a representation of the interactional space that is fundamentally different from the one we are used to in video-based analyses of talk-in-interaction. In the majority of this work, only one camera is used which records the scene from the perspective of a bystander, i.e. a hypothetical observer who watches from some distance (usually from an orthogonal position), without being an active participant. Due to the location and distance of the camera, most statements on eye-movement that can be derived using this method are no more than *reconstructions* of gaze on the basis of head movements, i.e. they rely on estimations and inferences rather than on the observation of actual gaze patterns. However, participants' gaze movements are by no means always accompanied by head movements; particularly the difference between gaze at a co-participant and gaze away from this co-participant are almost impossible to reconstruct in this way.¹ In addition, this camera perspective is not that of the co-

¹More sophisticated recording settings with more than one camera can overcome some of these methodological restrictions and approximate the accurateness of eye-tracking. Kendon (1967), in his pioneer work on gaze for turn-taking (see below), already abandoned the traditional perspective of a single camera 'observing' the scene from the outside in favor of a setting in which the camera filmed one of the two participants over the shoulders of the other, with a mirror placed next to the directly filmed participant

participants, which Conversation Analysis usually considers to be central. In the reality of interaction, co-participants each have a different field of vision, and hence a different view of the others.

Of course, it would be naïve to believe that scene cameras with the superimposed gaze tracker provide a 'real' representation of coparticipants' vision, let alone of their cognitive representations of the interactional space and the world around them. For instance, the width of the scene camera is considerably narrower than natural vision, at least for people who do not wear glasses with heavy rims. Natural mid-peripheral vision extends to 60° on both sides, far-peripheral vision even to up to 120°, while the scene cameras of eyetracking systems usually extend only 90° horizontally and 50° vertically. Also, the tracker does not allow us to identify the participants' focus of attention, but only the direction of their gaze (cf. research on "mindless reading", e.g. Reichle, Reineberg & Schooler 2010). An absent-minded person may appear to be 'looking at' something (i.e. the tracking point is fixed on this object) but is in reality not paying attention and perhaps not even perceiving this object. However, what we can identify with a considerable degree of accuracy is the direction of participants' central vision (gaze), which is also what co-participants can see.

3. Turn-taking and gaze

The celebrated 1974 article by Sacks, Schegloff and Jefferson on turn-taking is rightfully considered one of the foundational texts of conversation analysis. Inspired by their work, there has been a continuous and ongoing interest in turn design and the projectability of turn completion in conversation analysis, including recent studies on multimodal turn

showing the face of this second participant. Multiple synchronized cameras can be used to capture multi-party settings, but this is rarely done in conversation analysis as they are considered to be too obtrusive and the data obtained in such a way 'unnaturalistic'.

packages, the role of prosody and syntax, but also turn expansions (cf. the overviews in Couper-Kuhlen & Selting 2018: Ch.2, Clift 2017: Ch. 4) and collaborative turn constructions (cf. Hayashi 2013). There is also a large body of research on how participants display reciprocity (among other things, by gaze, see Goodwin 1980, Holler & Kendrick 2015), on collaboration between the speaker and the recipient in the emergence of turns (Auer & Zima, submitted), on overlaps and interruptions (see, e.g. Schegloff 2000), and even on universals of turn-taking across cultures (Rossano, Brown & Levinson 2009). However, the core of the model, i.e. what Sacks and colleagues labelled the “turn allocation component”, although quoted in all introductions, has received far less attention.

As is well known, the “turn allocation component” has three “sub-rules” (SSJ 1974: 703-4), the first (IIa) of which gives the current speaker the right to select a next speaker; the second sub-rule (IIb) stipulates that if the current speaker does not exert this right, then any of the other participants may self-select, with the “first starter” acquiring the right to the turn. Finally, if neither the current speaker selects a next speaker, nor any other participant self-selects, sub-rule IIc applies, and the current speaker may continue. Reminiscent of rule ordering in Generative Grammar of the time, the three sub-rules are presented in terms of a hierarchical relationship: (a) applies before (b) and (c) only after (b) has applied. The full relevance of the model can only be investigated in multiparty interaction (hence the reliance on three party constellations in this paper), as the next speaker in two-party interactions is always the other speaker, hence (a) and (b) are irrelevant.

The “current speaker selects next”-techniques are of course a crucial issue in this model, which the first sub-rule refers to. Here, Sacks and colleagues make a statement which massively reduces the applicability of the first sub-rule: “[t]he group of allocation techniques which we have called ‘current speaker selects next’ cannot be used in just any utterance or

utterance-type whatsoever. Rather, there is a set of utterance-types, adjacency pair first parts, that accomplish such selection[.]” (1974: 710-11) Only when formulating a first pair part can the speaker choose a co-participant as the next speaker who then has the right and obligation to formulate a second pair part. Even though Sacks et al. mention the possibility for turns to be retrospectively transformed into first pair parts by a post-positioned question tag, this restriction of sub-rule IIa to first pair parts means that after a large group of actions by a current speaker, self-selection is the only option.

This distinction between one group of actions that have the status of “firsts”, and which need to be responded to with a defined second action in the adjacent position, and another group of actions for which no next actions are projected, has been criticized by Stivers & Rossano (2010) who argued that it should be replaced by a continuum of more or less projecting first actions. Moderately projecting first actions would be, for instance, first assessments or tellings, while strongly projecting first actions are questions or invitations. I will follow this revised view in this paper, but will go beyond it by arguing that the current speaker can select a next speaker even in non-projecting contexts, particularly through gaze.

There are sequential contexts mentioned by Sacks and colleagues in which explicit current-speaker-selects-next techniques are not necessary, as the next speaker is already determined by sequence structure. For instance, repair initiations or confirmation requests will always select the participant whose turn is in the scope of the repair or confirmation as the next speaker.

If first speakers need to allocate the turn to a next speaker by specific techniques (i.e., after a first pair part), the main technique Sacks et al. seem to have in mind is the use of names and other referring expressions such as address terms. In addition, they refer to “techniques which employ social identities in their operation” (1974: 718), which can be understood as a

special case of recipient design based on social categorization. Gaze is mentioned in passing only,² even though at the time, a number of film-based interaction studies on turn-taking had already been published, most notably Kendon's ground-breaking paper on the role of gaze for turn transition (published in 1967).

Address terms (such as names or kinship terms) and gaze function very differently in terms of turn-allocation: non-pronominal address terms are a very forceful, explicit way to select a next speaker, but they occur only rarely in this function, at least in my (German) data. In fact, they "appear to be deployed to do more than simply specify whom the speaker is addressing", as Lerner (2003: 184) notes. As they also occur in two-party interaction, their main function seems not to be allocation of turns, but to appeal to the addressed person, or to express the speaker's affection for him/her (Günthner 2019, 2016; Droste & Günthner 2000; Clayman 2010, 2012; among others). Gaze, in contrast, is available as a resource for turn-allocation at all points in face-to-face interaction provided that participants are able to see one another; it is independent of the verbal action performed by the current speaker. It can even be argued that it is difficult in F-constellations to *not* use gaze for next speaker selection in Western cultures, since looking away at the possible end of a turn is a turn-holding technique.

As Lerner (2003) points out, gaze is unreliable in the case of non-attentive co-participants, or those whose visual attention is devoted to other interactional tasks. It is the combined use of gaze and second person singular pronouns which can remedy this disadvantage. While the pronoun *you* itself is not able to select a co-participant as next speaker, it provides a verbal cue which attracts coparticipants' visual attention to the speaker's gaze which then, in turn,

²"Thus, an important, perhaps the central, general technique whereby current speaker selects next, involves the affiliating of an address term (or some other device for achieving 'addressing', e.g., gaze direction) to a first pair-part." (SSJ 1974: 717).

selects the gazed-at participant as the next speaker. Hence, according to Lerner, gaze in combination with the second person pronoun is the prototypical and best current-speaker-selects-next technique in multiparty interaction. But this, of course, does not mean that gaze alone cannot be efficient for selecting a next speaker as well.³

Kendon (1967) investigated gaze systematically as a turn-allocation technique for the first time and arrived at one of the most fundamental generalizations: it is the recipient who needs to look at the speaker much more than the speaker who needs to look at the recipient. However, his study was based on two-party interactional data, so that techniques of next speaker selection remained largely beyond his interests. According to his results, turn transition in two-party interaction relies on cues in the prospective next speaker (who looks away before starting an elaborate next turn, while he keeps looking at the speaker when he wants to remain in the participation role of the recipient) and in the current speaker (who tends to look at the prospective next speaker when he intends to finish his turn). Goodwin (1981, etc.) extended these findings to multi-party interaction and found that prospective speakers have to secure the gaze of at least one recipient before they can take the turn. The imbalance between recipient's gaze at the speaker and the speaker's license to look away from the recipient is also indirectly confirmed by the way in which overlapping talk is resolved; here, the successful competitor for the floor tends to look away, i.e. to display a gaze pattern which is typical of a speaker, not of a recipient (see Zima, Weiß & Brône 2019).

Also working on two-party interaction, Rossano (2012) uncovered another gaze pattern according to which speaker and recipient sustain gaze over a sequence of (usually short or sequentially tightly organized) turns, while they avert gaze when the sequence is terminated.

³As Lerner also points out, there are other, more implicit ways of selecting a next speaker. Most efficiently, turns can be tailored to the specific epistemic status of a particular co-participant in a way that makes only this participant eligible for the next turn.

This pattern suggests that the type of sequence (i.e. the actions performed) may have an impact on the co-participants' gaze behavior. Rossano (2006) and Stivers & Rossano (2010) additionally found that prolonged gaze can be used to pursue a response when a silence occurs after a sequentially implicative first activity and the response does not seem to be forthcoming.

One of the few papers in which the regulatory functions of gaze in multiparty interaction were studied is Tiitinen & Ruusuvuori (2012) who investigated interactions between two parents and a nurse in a maternity clinic. In these interactions, the nurse was observed to gaze at the addressed participant after a question that was verbally addressed at both parents by use of the second person plural pronoun; it was significantly more often this gazed-at parent who answered the question. Vranjes (2018, Ch. 4) studied gaze in triadic interpreting situations and found that the precise onset of the interpreting activity is negotiated on the basis of gaze by the current speaker.

A comprehensive study on gaze and turn-taking in three-party interaction was conducted by Weiß (2018, 2019, 2020). She focuses on cases in which the gaze-selected person does *not* follow the gaze-selection of the current speaker, i.e. he or she deviates from the expected pattern. For instance, a gaze-selected participant can 'refuse' to take the turn and instead gaze-select a third participant thereby transferring the turn to this participant. Weiß shows that

- current speakers who are confronted with a co-participant who, although gaze-selected, does not take the turn, may continue to gaze at this pre-selected participant even beyond the completion of their TCU, i.e. into a post-completion pause, thereby prompting a response;

- gaze-selected next speakers may signal their rejection of the turn offer by looking away from the current speaker *during the last part of his/her ongoing turn*, thereby symbolically withdrawing from the interaction and making themselves unavailable as next speaker; this contrasts with next speakers who look away *during the beginning* of their turn (Kendon 1967);
- gaze-selected next speakers may, instead of taking the turn, gaze-select the third speaker and thereby 'hand over' the turn to this participant.

There is also a small number of experimental studies on turn-taking and gaze. An early contribution by Kalma (1992) is based on elicited data from 40 discussion groups of three students each. He identified a gaze pattern in "high-influence" speakers which consists of looking at another speaker at the end of the turn and "prolonging" their gaze into a pause after they had stopped speaking. Kalma argues that these speakers "seem to be inviting the person they are looking at to take over", which occurred in 95% of all cases (Kalma 1992:29). This coincides with Stivers & Rossano's claim that gaze is one of the parameters that "mobilize" a (gazed-at) participant's response (Stivers & Rossano 2010). For instance, Stivers & Rossano found that of 336 requests for information in Italian and English conversation, 61% were accompanied by speaker's gaze at the addressed participant who was selected to deal with the request. The response-mobilizing force of gaze is not only observed after first pair parts, but also in weakly projecting contexts.

In sum, research on gaze and next speaker selection so far has accumulated evidence that the gaze of the current speaker is a technique to allocate the turn to the gazed-at participant. However, existing research has either been restricted to the sequential context of adjacency pairs, i.e. next speaker selection in a strongly projecting environment, or it is based on dyadic interactions, in which turn transition, but not next speaker selection, is an

issue. In addition, most studies are based on video-recordings with the above-mentioned restrictions on observational accuracy, or on experimental data that reduced ecological validity. Only a small number of studies (particularly those by Weiß and Vranjes) use eye-tracking technology in spontaneous, non-experimental settings.

4. Data and methods

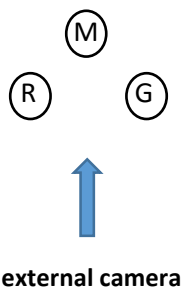
The data used for this study (the same which Weiß used in the study on deviant cases) come from three-party interactions in groups of students sitting around a table, who were recorded with SMI eye tracking glasses (in addition to an external video camera). Note that in this constellation, gaze is not needed for movements or object-manipulations in space, as for instance, when participants are walking together, which requires the use of gaze for path-finding, spatial orientation, checking other walkers, etc., or when they are engaged in bodily activities (cooking, eating, etc.) (see Stukenbrock 2018a & b, *forthc.*, Stukenbrock & Dao 2019 for the use of eye-tracking in these contexts). In these cases, gaze has to deal with many competing tasks which obviously complicates analysis.

For the present paper, six recordings of roughly 60 minutes' duration each were used.

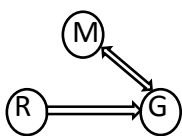
Analysis was performed using ELAN, the verbal transcription follows GAT2 (Selting et al. 2011). A simple, iconic transcription system for gaze was developed which is based on Rossano (2012) and was further developed in Auer (2018). It shows the co-participants' gaze from a bird's eye perspective. The system distinguishes between gaze at another participant and gaze away from the co-participants (at an object or into open space). Gazing 'at' another participant is operationalized as a fixation of this participant's gaze in the facial region of the

gazed-at person.⁴ In the first case (gaze at co-participant), which is most relevant for this study, double arrows are used that represent the gaze of each speaker. In the second case (gaze away from co-participants), single lined arrows are used. If a participant looks down, the arrow has a broken line.

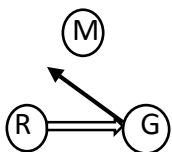
For instance, the three speakers shown in Fig. (1) with the pseudonyms G(eorg), right, M(arion), middle, and R(asmus) are seated like this:



with the external camera facing Marion. At that particular moment, Rasmus (whose view is shown on the left lower picture) is gazing at Georg, Georg (right upper picture) at Marion, and Marion (left upper picture) at Georg:

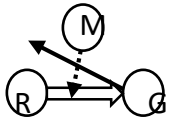


If instead of looking at Marion, Georg would have directed his gaze into the space between R and M, this would be marked by a simple arrow:



⁴Given the distance of 1.5 to 2 meters between the participants, this is a realistic “resolution” which may come close to participants’ interactionally relevant discrimination. Humans closer to each other may be able to monitor more subtle eye movements.

And if Marion had also looked away from both co-participants, and down, this would be marked by a broken arrow:



In the following analysis I want to develop the hypothesis *that the speaker's gaze in the last part of the ongoing utterance, before a possible completion point is reached, is a technique of selecting a next speaker.*⁵ In the following, only gaze at a co-participant will be considered that starts at least 0.4 seconds before the possible completion point of this turn.

The speaker may (and usually will) look at different co-participants during his or her turn in multi-party interaction. Gaze alternation for addressee selection during longer turns is a solution to an interactional problem inherent in gaze: a speaker can only look at one co-participant at a time (in the spatial constellation we are dealing with here). By alternately looking at the addressees, current speakers actively keep a three-party conversation from turning into a two-party conversation by gaze-addressing and gaze-engaging both co-participants (Auer 2018). It is only in the last part of the turn that gaze assumes turn-taking relevance. During a short turn, the functions of gaze for selecting addressees and for selecting a next speaker may coincide. In this case, the addressee of the turn and the selected next speaker are identical.

5. A typology of turn-allocation techniques and their sequential embedding

⁵ How long exactly this 'last part' has to be, still needs to be investigated and will not be the topic of this paper. But the duration of gaze must be sufficiently long for potential next speakers who may not have planned to produce a next turn before the current speaker's gaze has reached them to start producing their turn without a noticeable gap. Given an average lapse of 0.2 sec in unmarked turn-taking (Rossano, Brown & Levinson 2009), and a minimum planning time of no less than 0.6 sec (Levinson 2016) it would seem unlikely that a period of less than 0.4 sec would be sufficient for such a smooth turn start.

In this section, I will develop a comprehensive scheme for the analysis of turn-transition in three-party interaction which covers all sequential contexts in which next speaker selection is an issue around a (possible) turn completion.

For such a comprehensive approach, two dimensions of turn-by-turn talk are necessary: the **dimension of sequentiality** (i.e. the projection of a next action on the basis of the sequentially preceding action) and the **dimension of next-speaker selection**. Understanding sequential projection as a gradual dimension means that current-speaker-selects-next techniques also apply to cases in which the projection of a next action is weaker than in an adjacency pair. Understanding next speaker selection as a gradual dimension means that the model needs to do justice to the variety of turn-allocation techniques that are used in interaction and which include gaze as the most ubiquitous. These turn-allocation techniques may be of different strengths as well, and they can be combined with each other. This results in a two-dimensional space (see Fig. (2)).

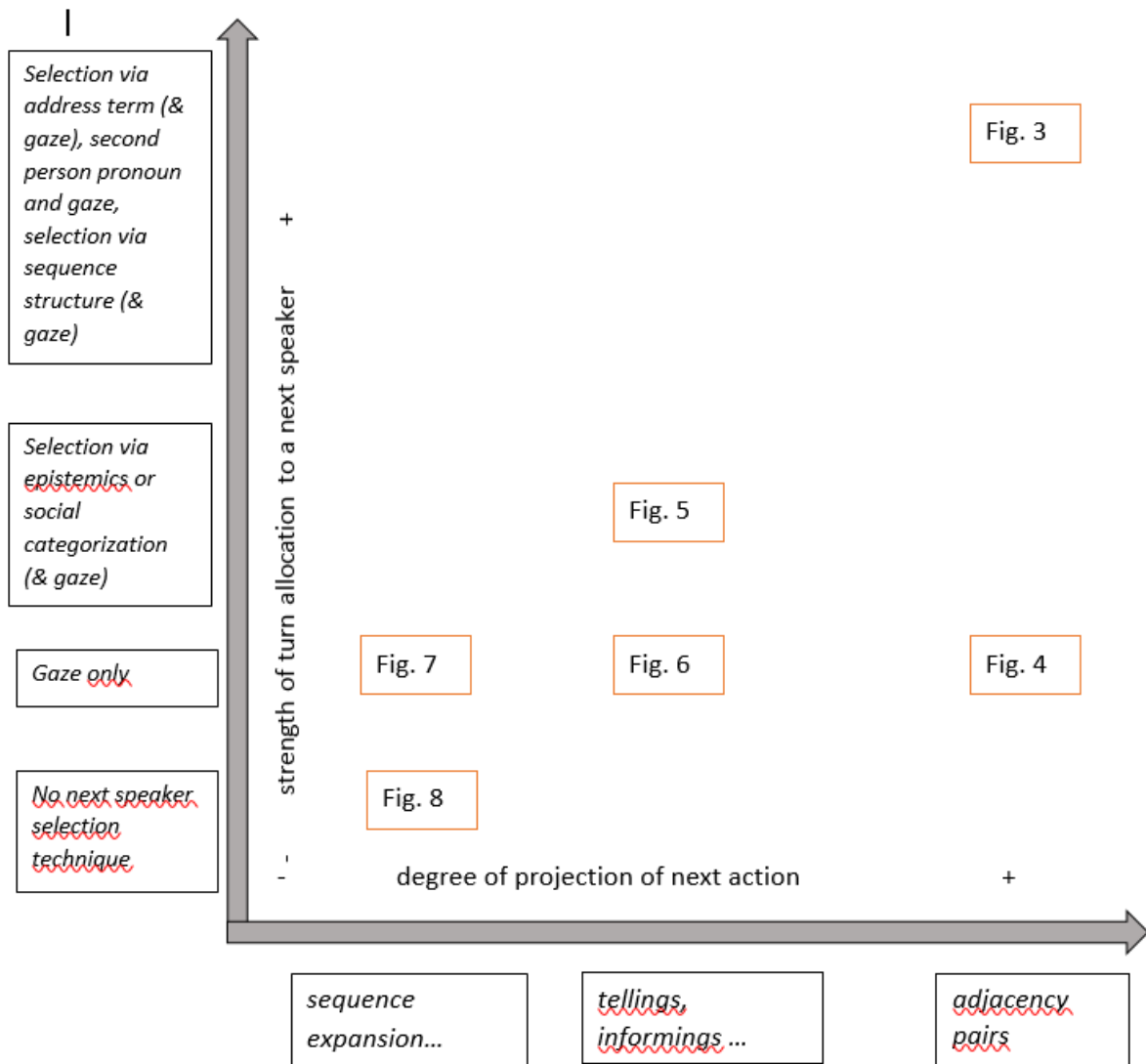


Fig. (2): The 'degree of projection' dimension and the 'strength of next speaker allocation' dimension spelled out for individual points on the continua. Numbers refer to the data extracts discussed below.

On the horizontal axis, the proposed continuum ranges from maximally projecting first actions such as adjacency pairs, to actions that do not project any next action, such as the ones that are typically found after sequence closure; whatever happens here is either a sequence expansion or the beginning of a new sequence. In between, we find weakly projecting actions such as tellings or informings, which require some kind of next activity by the recipient but do not specify this activity precisely (a telling may be followed by an acknowledgement, an assessment, a second telling, etc.).

On the vertical axis, the continuum is more hypothetical, but ranges from cases in which the current speaker selects a next speaker through unique selectors such as address terms or second person pronouns together with gaze, down to next speaker selection based on gaze alone. Intermediate may be allocation techniques such as a specific recipient design, based on epistemic and social grounds, again in possible combination with gaze. The minus pole of this dimension is represented by cases in which no technique of next turn allocation is used at all and in which self-selection can take place.

The following data extracts may serve to illustrate the different possibilities. Let us begin with the somewhat trivial case of turn-taking after a first action that strongly projects a second, and in which the current speaker also uses strong next-speaker selection techniques, here the second person singular pronoun in combination with gaze.

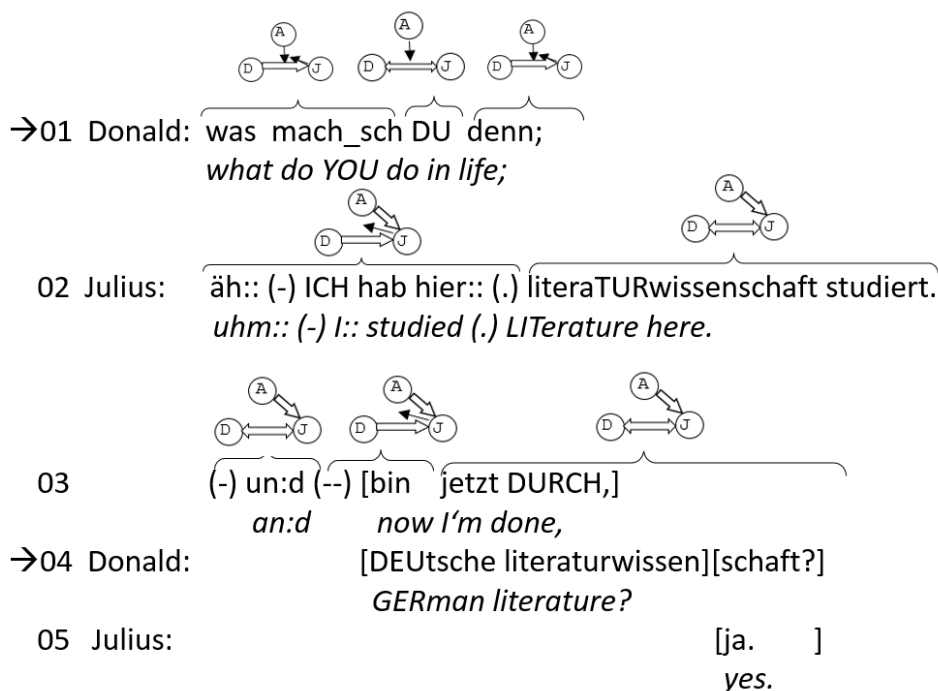


Fig. (3) strong projection, use of 2nd sg. pronoun and gaze for next speaker selection techniques

We are in the early phase of the episode in which the three students introduce themselves. In the beginning of the extract (line 01), Donald (sitting left) asks Julius (sitting right) about his studies. On the dimension of next action projection, a question for information strongly projects an answer in the adjacent slot. Julius provides such an answer by saying that he studied literature (line 02), and after a little silence, expands his turn to add that he has finished in the meantime (line 03). In overlap with this expansion, Donald asks a follow-up question in the format of a confirmation request for his guess that Julius studied *German* literature (line 04), which Julius confirms. This closes the sequence.

With regard to the dimension of next speaker selection, the second person pronoun *du* in the sequence-initial question is part of a multimodal (deictic) package in which the combination of gaze and the pronoun uniquely selects Julius as the next speaker. Julius complies (second line), while Alma, the third participant in this interaction, remains silent.⁶ It is this pattern that Sacks and colleagues had in mind in the formulation of sub-rule 2a in the turn-allocation component of their model.⁷

In extract 2, an equally strongly projecting first action (again a question) is produced in such a way that it is verbally addressed to both co-participants by the use of the second person plural pronoun, German *ihr* 'you [PL]'. In this case, only gaze indicates who the speaker wants to select as the answerer:

⁶ For the follow-up question, Donald keeps his gaze on Julius, thereby selecting him again as the next speaker. Although the confirmation request "DEUtsche literaturwissenschaft?" does not contain a second person pronoun, its format nonetheless uniquely selects the previous speaker, Julius. Next-speaker selection here has a sequential basis (it is a follow-up question which is addressed at the same co-participant as the first question), but additionally combines with TCU-final gaze at the recipient. Again, Julius takes the turn in overlap with Donald's follow-up question and confirms (see last line).

⁷ Donald always looks at Julius. Julius, on the other hand, looks away from Donald, the current speaker, during the first phase of Donald's initial question, then focusses on Donald for a short period (one word: "du"), and then withdraws gaze from him again – a typical pattern for a next speaker already described in Kendon (1967).

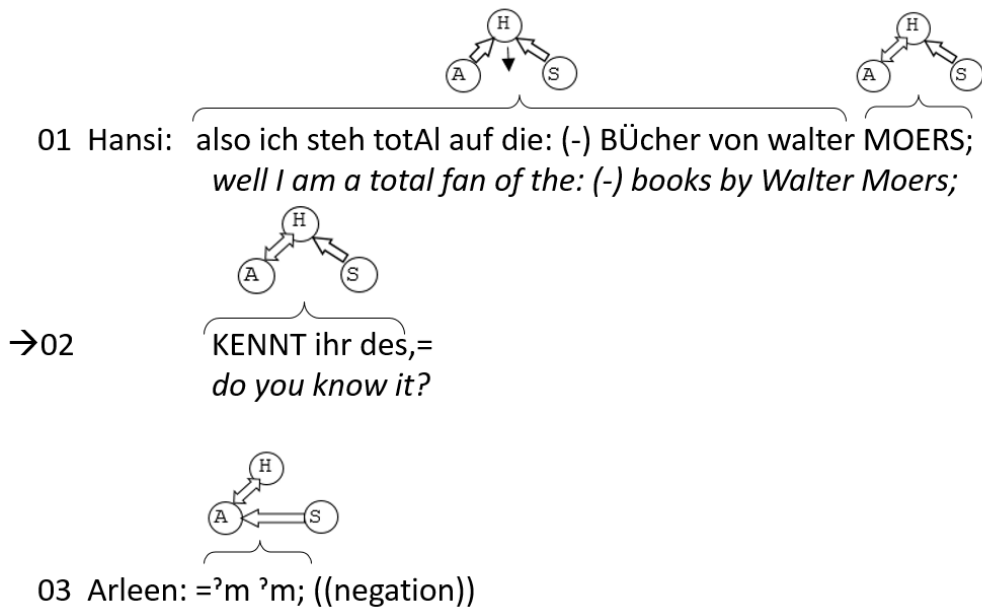


Fig. (4): strong projection, speaker selection by gaze only

Hansi, the speaker sitting in the middle, talks about her favorite writers. She mentions an author and asks the others whether they know his books (line 01). The question addresses both co-participants, Arleen and Svetlana, verbally via the second person plural pronoun. However, it is only one of them – Arleen, sitting on Hansi’s right– who answers by saying that she does not (line 03). Why is it Arleen who answers? The selection of Arleen is exclusively based on the questioner’s gaze. During the production of her turn-initial statement in line 01 (the assessment of the writer *Walter Moers*), Hansi mostly looks into the open space in front of her. It is only when she reaches the last word of this turn constructional unit (the proper name *Moers*), which also carries the nuclear accent of the intonation phrase, that she turns her gaze toward Arleen, and thereby selects her as the next speaker. She keeps looking at Arleen during the second part of her turn, i.e. the question ‘do you know it?’ (meaning: his book, line 02). Arleen, who has been looking at the speaker throughout, takes up this cue and answers immediately, while Svetlana remains silent (and by looking at the answerer accepts Arleen’s claim to the turn).

We now turn to less tightly organized sequences, i.e. those in which the first action projects a particular next action in a weaker way only. In extract (3), this first action is an evaluative statement that compares the educational systems of two German states. Participants are again Alma, Julius and Donald.

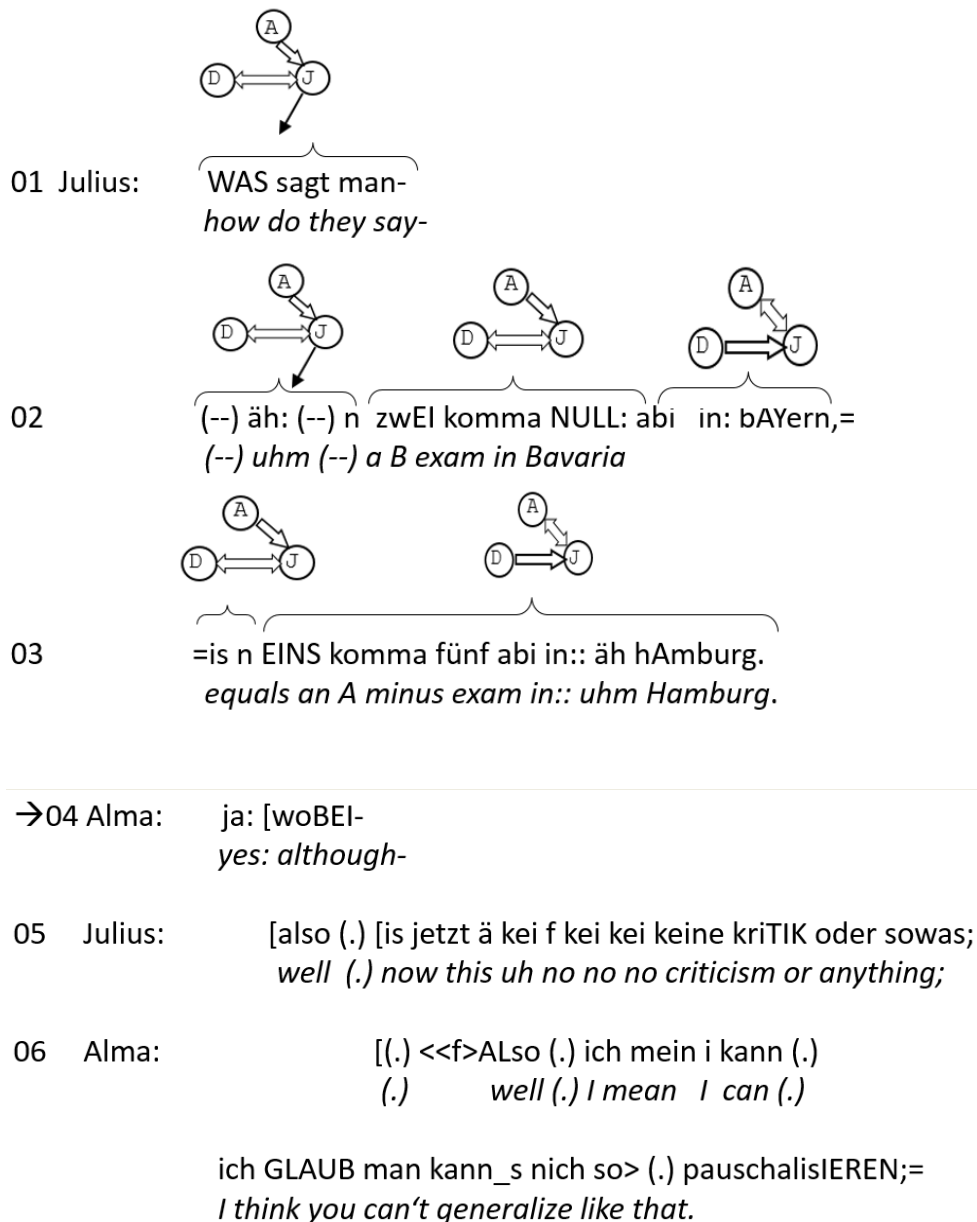


Fig. (5): Moderately strong projecting first action and next-speaker selection by social categorization and gaze

The main speaker in this sequence formulates a stereotype (cf. his initial framing: 'how do they say') in Germany according to which the northern states such as Hamburg give much better grades for the same school achievements than the southern states, such as Bavaria. What makes this turn somewhat delicate is the fact that shortly before, Alma had told the others that she grew up in Hamburg. Julius can therefore be considered to be threatening Alma's face by downgrading the quality of her school exams, even though this is not made explicit. A response by one or both co-participants that expresses their opinion on the topic is projected, but the absence of such a second action (for instance due to the co-participants' ignorance or indifference vis-à-vis this matter) would not be as sanctionable as the absence of an answer after the questions in extracts (1) and (2). If Donald's sequence-initial action is heard as a critique, a response is of course even more projected.

During his comparative evaluation, Julius looks alternately at Donald and Alma, addressing them both. But based on Alma's self-categorization as a 'person from Hamburg' (who also attended school there), she is more likely to comment on Julius's statement, particularly if she feels the need to defend herself against the imputation of a final exam without merits.

In addition to issues of social categorization (Alma as a 'northerner'), next-speaker-selection is also achieved by Julius's gaze. In the beginning of his turn he looks away into the open space before him (line 01 and beginning of line 02), a typical gaze behavior during hesitations, presumably linked to planning activities. Once the hesitation phase is over (on "zwei komma null", 'B-exam', line 02), he first turns to Donald and then to Alma, just to turn back to Donald for a very short moment (on "in_n", 'is a') in the beginning of the predicate. But when approaching the end of his turn (line 03), he gazes at Alma, selecting her as the next speaker. And indeed she picks up the topic eagerly and starts to disagree, arguing that things should not be generalized (lines 04/06).

Even further down on the dimension of next-speaker selection strength are cases in which no other technique than gaze is employed. In these cases, the first action can still be sequentially projecting; nevertheless, it is verbally designed in such a way that it can be responded to by all co-participants.

The following sequence is again taken from the very beginning of the episode between Alma, Julius and Donald, immediately after the initial exchange of greetings and self-presentations. Donald starts the conversation with a kind of excuse ('I thought I might arrive a few minutes late but...', line 01-02), which is ritualistic as he had in fact arrived in time. A ritual excuse need not, but can be responded to by acceptance. Both co-participants, Alma and Julius, are eligible for this next action.

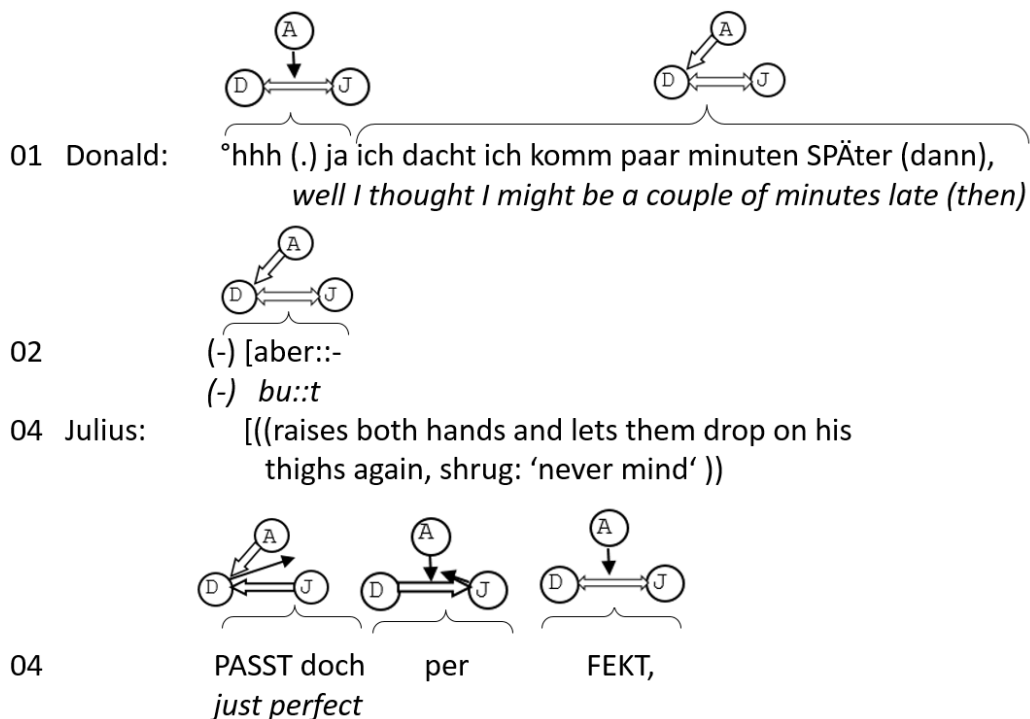


Fig. (6): Moderately projecting first action and next-speaker selection by gaze only

However, Donald selects Julius as the preferred next speaker by looking at him, and Julius also provides a fitted second action by accepting the apologies in line 04.

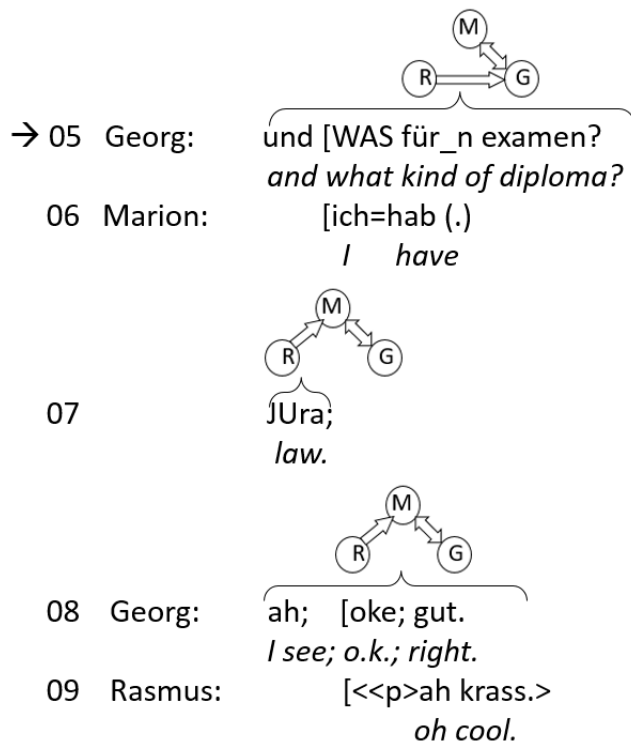


Fig (7): Absence of action projection and next-speaker selection by sequence and gaze

Marion corrects Georg's inference that she will receive her master's degree by saying that she will receive a "(state) diploma" (line 02), and Georg accepts this correction in line 03. At this point, the sequence is closed; Georg's question has been answered, and gaze aversion is expected (Rossano 2012 for two-party interaction) (see arrow). But Marion and Georg keep looking at each other. Marion's sustained gaze invites Georg to expand the sequence as well as the topic with a follow-up question, by asking Marion (in line 05) about the kind of diploma she will receive (see Stivers & Rossano 2012). She answers that it will be a law diploma which leads to positive receipts by Georg (line 08) and Rasmus (line 09). These receipts close the sequence again.

As Marion and Georg have been in a colloquy, a continuation of this colloquy is more likely than a sequence expansion by the third participant. Hence, on sequential grounds, Georg is an expected next speaker. In addition, however, gaze selects him for this role.

6. A quantitative check

The main claim of this paper is that gaze is an efficient and ubiquitous way of selecting next speakers, co-occurring with other techniques of turn-allocation or alone. Gaze can be used regardless of the projecting strength of the verbal action involved.

This hypothesis was tested in a small quantitative study with the aim of finding out

(1) whether the direction of gaze during the final part of the turn can predict who will take the turn next in three-party interaction and

(2) whether gaze is more efficient when it is combined with other turn-allocation techniques (as predicted by the dimension of next-speaker selection strength in Fig. (3)).

The study was based on 250 turn-transitions taken at random (but in an uninterrupted sequence) from four different three-party interactions in the corpus (i.e., based on 12 speakers). Seventeen examples had to be discarded for various technical reasons, which means that 233 cases could be used. Cases of self-selection (i.e. those in which no potential current-speaker-selects-next techniques were observed) were not included. Utterances merely displaying the participant's reciprocity ("continuers", etc.) were disregarded as well since they do not constitute cases of turn-transition. The coding (performed by four different trained collaborators and cross-checked, but without a formal cross-coder agreement test) included the following points on the (horizontal) dimension of action projection, based on the action taken by the second speaker:

- strong projection (the action by the second speaker is strongly projected by the action of the first speaker)

- moderate projection (the action by the second speaker is moderately projected by the action of the first speaker)
- no projection (the second speaker either expands the sequence or starts a new sequence).

On the horizontal dimension (next speaker selection), the following techniques as used by the first speaker were coded.

- Maximal cues: This included (a) the use of the 2nd person singular pronoun together with gaze at the selected participant, (b) sequence-based next speaker selection with or without gaze selection, (c) names or other address terms with or without gaze selection, (d) combinations of the above. In one case, the first speaker used pointing at the selected next speaker, which was also included in this group.
- Next speaker selection based on epistemics and/or social categorization with or without gaze selection, but without any of the cues listed in the first category.
- Only gaze.

There was not a single case in the data in which a verbal next speaker selection technique was used without gaze; this is not to say that they are impossible, but it seems that this is a highly marked case in a seating arrangement such as the present one in which the participants are always visible to each other. In the category of maximal cues, it deserves to be mentioned that the data set included only one name in the function of an address term. The 2nd person singular pronoun occurred in six cases. Hence, the main next speaker selection techniques observed in this data set were based on sequence, social categorization and/or epistemics.

Table (1) summarizes the results (raw data). Cases, in which turn taking was accompanied by turn-taking disturbances⁹ were counted as counterevidence (=third participant takes over). The number of turns that are produced without being sequentially projected is high in this corpus (164 out of 233 instances, i.e. 70% of all cases). The proportion of cases in which the speakers in these turns were selected by gaze only is also quite high (88 of 164 instances, i.e. 54%, vs. 34% - 55 out of 164 – in the case of maximal speaker selection cues). If we compare them to the group of strongly projected next turns, we can see that in their case, there is a very marked tendency to select the next speaker with strong ('maximal') cues

	<i>Strong projection</i>		<i>Moderate projection</i>		<i>No projection</i>	
	<i>Selected speaker takes over</i>	<i>Third participant takes over</i>	<i>Selected speaker takes over</i>	<i>Third participant takes over</i>	<i>Selected speaker takes over</i>	<i>Third participant takes over</i>
<i>Maximal cues</i>	34	3	10	3	50	5
<i>Epistemics/social categorization as cue</i>	0	0	0	0	21	0
<i>Gaze only as cue</i>	5	3	11	0	69	19

Table (1): Number of turn transitions to the participant selected by the previous speaker vs. number of turn transitions to third (non-selected) participant, in relation to projective strength of first action and cues used for next speaker selection.

(37 out of 45, i.e. 82%). Only rarely (8 cases) are next speakers selected by gaze alone in these cases. The ways in which a next speaker is selected and the projection of the turn for which this speaker is selected are therefore related. However, the interdependency is stronger in the case of strongly projected actions and 'maximal' selection cues (82% vs. 18%) than in the case of non-projected next actions and 'gaze only' as a next speaker selection technique (54% vs. 34%). The correlation between projective force and the type of cues for

⁹ This was the case in 14 transitions.

next speaker selection is highly significant (Yates chi-square 24.51, df. 1, $p < .0001$; Cramer's $V = 0.3735$).

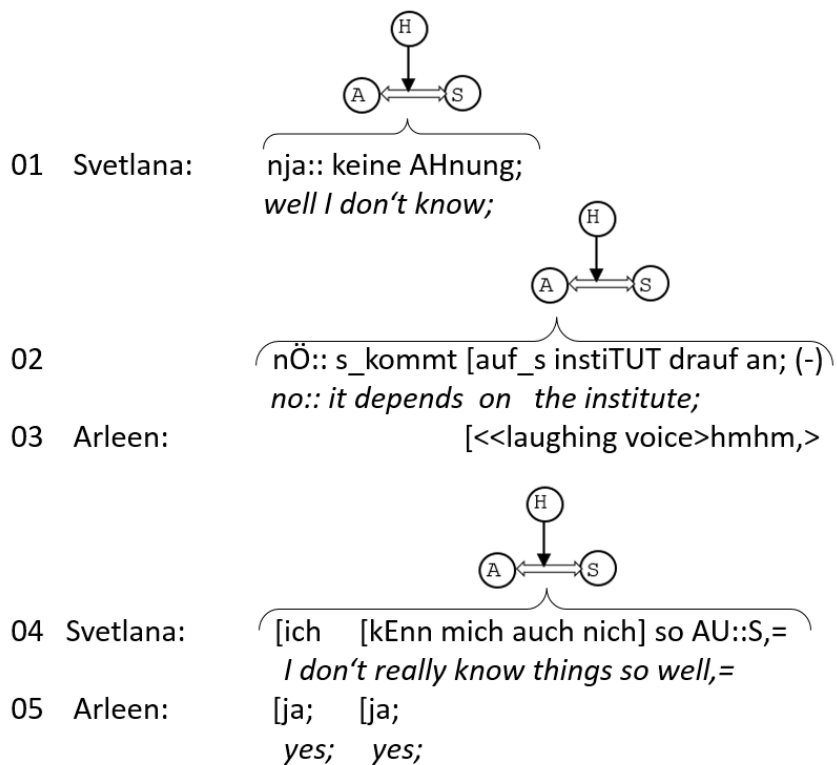
Selection of next speakers via their social categorization or background knowledge (epistemics) is relatively rare in this corpus (9% or 21 cases), and it always occurs in non-projected turns, although this result should not be overinterpreted given the small total number of transitions.

But the main quantitative result is of course that in the overwhelming majority of turn transitions it is the participant who is gazed at during the final stretch of talk of the previous speaker who takes the turn (86%, or 200 of all cases), confirming hypothesis (1). This is generally true, regardless of the strength of projection or the number of other speaker-selecting cues. Yet there are differences: if next-speaker selection is based on gaze only, the number of 'mismatches' rises to 20% (22 out of 109 cases), while in the case of maximal next speaker selection cues (which always include gaze), it is somewhat lower (94 vs. 11 instances, i.e. 10%). The difference not significant, however. It remains therefore an open question whether gaze in isolation is less efficient in selecting a next speaker than verbal techniques combined with gaze, as predicted by hypothesis (2).

7. Discussion

So far, I have shown that conversationalists' use of gaze as a next-speaker selection technique is extremely widespread and that gaze is a highly efficient way of allocating turns. Its use extends much beyond what Sacks and colleagues seem to have had in mind when discussing current-speaker-selects-next techniques. Including gaze among the techniques for next-speaker selection expands the applicability of their "sub-rule 1a" greatly. Obviously, this has consequences for "sub-rule 1b" as well, i.e. for self-selection. Self-selection only applies

in the rare cases in which not even gaze selects a next speaker, for instance, when the current speaker looks away from all co-participants at the end of the turn. The turn is then free to be taken by any other participant. Often, this looking-away also suggests sequence and even topic closure so that self-selection as speaker coincides with the initiation of a new sequence and topic. Here is an example.



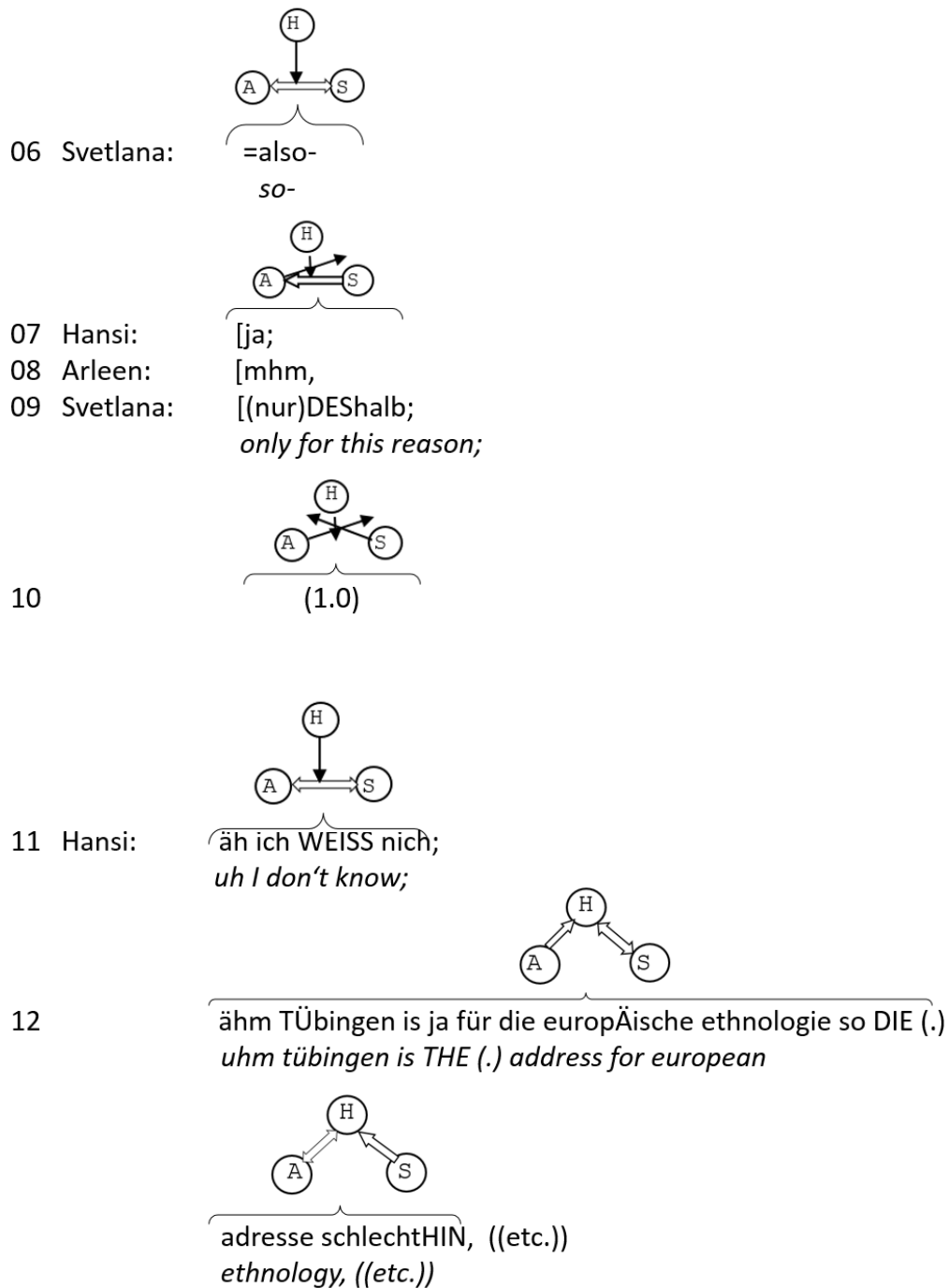


Fig. 8: Self-selection

Hansi (in the middle), Arleen (to the left) and Svetlana (to the right) have been talking for some while about Svetlana's plans to enroll at a university for a program in European Ethnology. Svetlana closes the topic by "well I don't know" in line 01, and by pointing out that where she wants to go will depend on the institute and that she does not really have an

overview (lines 02, 04). Arleen (05) and Hansi (07) agree with her point of view. During the beginning of this phase, Svetlana mostly has eye contact with Arleen, while Hansi does not look at either of them (up to and including the *also* in the 6th line of the transcript). During her last “mhm”, Arleen looks away as well (line 08). In the ensuing one-second silence, the main speaker in this sequence, Svetlana, turns her gaze away from Arleen and looks into the open space between Arleen and Hansi, too (line 10). Hence, there is no eye contact between any of the participants. At this point, the turn is indeed completely free. It is Hansi who now self-selects and adds a comment on the Tübingen institute for European Ethnology, thereby expanding the sequence. On the level of gaze, this is a new start as well. After an initial phase of hesitation (on “äh ich WEISS nicht” ‘uhm I don’t know’, line 11), Hansi looks at Svetlana (first part of line 12) and later Arleen (second part of line 12), and both Svetlana and Arleen turn to her, acknowledging her new speaker-role and accepting for themselves the role of recipients.

The proposed revision of the model also has consequences for the so-called *first starter principle*, cf. the following example from Sacks’ and colleagues’ paper:

(from Sacks, Schegloff & Jefferson 1974, example 1-3, data source: Labov: Battersea: A:7)

PARKY: *Oo what* they call them dogs that pull the sleighs.

(0.5)

PARKY: S-sledge dogs.

(0.7)

OLD MAN: Oh uh [:: uh

TOURIST: [Uh-Huskies.=

OLD MAN: Huskies. Mh,

[[

PARKY: *Huskies*. Yeh Huskies.

When sub-rule 1b (self-selection) applies, as in this case after Parky’s question in the beginning of this extract according to the traditional version of the turn-taking model, the

first starter will win. This, so Sacks and colleagues argue, is the reason why participants start a response turn with vocalizations such as hesitation markers even though they may not know what to say yet, which explains the old man's *oh uh::: uh* and the tourist's turn-initial *uh*, the latter being produced slightly after the old man's turn-initial vocalization. The motivation to start as early as possible when self-selection is possible is built into the turn-taking system. It accounts for the frequency of overlap. In the case documented by the transcript, the tourist answers the question and provides the lacking lexical item first (6th line), with the old man following a split-second later (beginning of 7th line of the transcript), even though both had equal rights to the turn.

But this is the picture without gaze. If we include it, it becomes clear that in most cases, it is likely that the two competitors for the turn do *not* have the same status: only one of them is the gaze-selected next speaker. Often it is the gazed-at participant who starts to speak first, with the non-selected participant being the second starter in the competition for the turn; and this is expected, since the first starter has selected this participant by gaze.¹⁰ If the inverse pattern is observed (the participant not selected by gaze is the first starter), the first starter must be seen as 'intruding' into the second starts turn-space.

In the following extract from our data, this can be demonstrated on the basis of the multimodal transcript. The topic is the moral scruples a criminal judge may have when "giving a defendant another chance" by means of a mild sentence.

¹⁰ For a multimodal analysis of overlapping talk including gaze also see Zima, Weiß & Brône 2019.

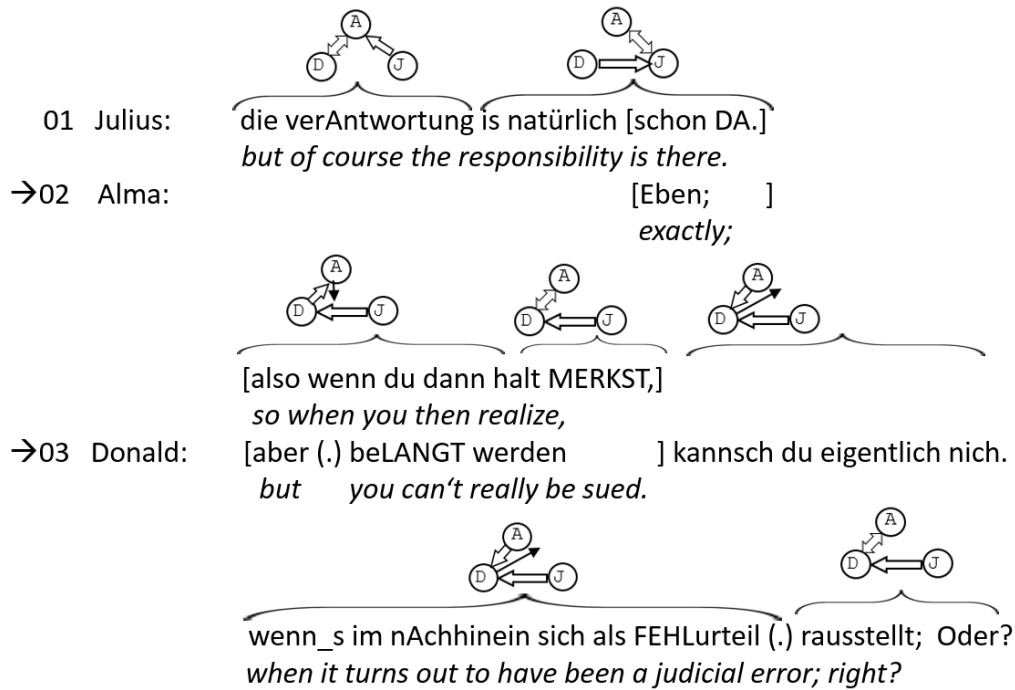


Fig. 9: Simultaneous starts

Julius argues in the first line of the transcript that the judge is responsible for the consequences of a mild verdict. In response to his statement, Alma and Donald compete for the turn; Alma is the first starter, beginning her agreeing turn two syllables before the end of Julius's turn. Donald is the second starter, who comes in a bit later and asks whether a judge can be sued for a judicial error. On the basis of the verbal transcript, Julius cannot be shown to have selected a next speaker and it appears that both incoming speakers seem to exploit the possibility of self-selection which is a consequence of his failure to select a next speaker. However, the multimodal transcript presents quite a different picture, as Julius clearly gazes at Alma and thereby selects her as next speaker. Alma's subsequent turn is therefore in accordance with the prior speaker's selection, while Donald's contribution is that of a non-

selected participant.¹¹ Neither is there self-selection, nor does the first-starter principle apply.

8. Conclusions

In this paper I have suggested to revisit one of the core feature of the turn-taking model, i.e. what Sacks, Schegloff & Jefferson (1974) labelled the “turn allocation component”, from the perspective of multimodal interactional analysis. I have argued that the inclusion of gaze into the analysis of turn-taking leads to a fundamental reappraisal of the role of self-selection for the turn. It greatly expands the domain of turn-allocation by the preceding speaker, as numerous cases of apparent self-selection of a next speaker turn out to represent invited speakership transitions. I have shown that gaze on the last part of a speaker’s turn at one of the co-participants in a multi-party setting is a very powerful way to select a next speaker. This holds regardless of sequential structure, i.e. in highly projecting sequential contexts just as well as after sequence closure. Among the various techniques that current speakers can use to select a next speaker, gaze is the most ubiquitous one (at least in the stationary three-party, talk-centered settings investigated in this paper). It usually accompanies and supports other techniques based on sequential structure, epistemics, or social categorization, and it may occur in tandem with second person pronouns, a particularly efficient multimodal package of next speaker-selection cues. However, even in the absence of these cues, gaze has been shown to lead to turn-transition to the gazed-at person (rather than the third participant) in the large majority of cases.

An exact documentation of gaze behavior requires the use of eye-tracking technologies, on which the present study is based. Progress toward a truly multimodal investigation of human

¹¹ It is a different question whether the gaze-selected co-starter also wins the competition for the turn. In fact, Weiß (ch. 5.2.) shows that this is not the case.

interaction can only be made if the limitations of traditional video-recording as it is still exclusively used in orthodox conversation analysis are overcome.

References

Auer, Peter 2018 Gaze, addressee selection and turn-taking in three-party interaction. In: Geert Brône & Bert Oben (eds.) *Eye-Tracking in Interaction. Studies on the Role of Eye Gaze in Dialogue*, Amsterdam: Benjamins, pp 197-231.

Auer, Peter & Zima, Elisabeth, submitted, Gaze, word search and co-constructions.

Brône, Geerd & Oben, Bert (ed.) 2018 *Eye-tracking in interaction. Studies on the role of eye gaze in dialogue*, Amsterdam: Benjamins.

Clayman, Steven E. 2010, Address terms in the service of other actions: The case of news interview talk. *Discourse & Communication* 4(2), 161-183.

Clayman Steven E. 2012 Address terms in the organization of turns at talk: The case of pivotal turn extensions, *Journal of Pragmatics*, vol. 44, no. 13, 1853–1867.

Clift, Rebecca 2017, *Conversation Analysis*. Cambridge: CUP.

Couper-Kuhlen, Elizabeth & Selting, Margret 2018, *Interactional Linguistics*. Cambridge: CUP.

Goodwin, Charles 1981, *Conversational Organization. Interaction between Speakers and Hearers*. New York etc.: Academic Press.

Droste, Pepe & Susanne Günthner 2020, „das machst du bestimmt AUCH du;“: Zum Zusammenspiel grammatischer, prosodischer und sequenzieller Aspekte syntaktisch desintegrierter du-Formate. In: Wolfgang Imo und Jens Lanwer (eds.), *Prosodie und Konstruktionsgrammatik*, 75–109. Berlin/Boston: de Gruyter.

Goodwin, Charles 1980, Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Social Inquiry* 50, 272-302.

Goodwin, Charles 2000, Action and embodiment within situated human interaction. *Journal of Pragmatics* 32, 1489-1522.

Günthner, Susanne 2019, Namentliche Anreden in onkologischen Aufklärungsgesprächen: eine interaktional ausgerichtete Studie zu Formen und Funktionen onymischer Anreden. Online: <http://arbeitspapiere.sprache-interaktion.de/82-susanne-guenthner-03-2019.pdf>. (retrieved May 26, 2020).

Hayashi, Makoto 2013, Turn allocation and turn sharing, in: J. Sidnell & T. Stivers (eds.), *The Handbook of Conversation Analysis*, Malden, MA: Wiley-Blackwell, 167-190.

Holler, J. & Kendrick, K. 2015 Unaddressed participants' gaze in multi-person interaction: Optimizing reciprocity. *Frontiers in Psychology*, 6: 98. doi:10.3389/fpsyg.2015.00098.

- Kalma, Akko 1992 Gazing in triads: a powerful signal in floor apportionment. *British Journal of Social Psychology* 31, 1, 21-39.
- Kendon, Almam 1967 Some functions of gaze direction in social interaction. *Acta Psychologica* 26, 22-63.
- Kidwell, Mardi 1997, Demonstrating reciprocity: knowledge displays as a resource for the unaddressed participant. *Issues in Applied Linguistics* 8, 2, 85-96.
- Lerner, Gene 2003 Selecting next speaker: The context sensitive operation of a context-free organization. *Language in Society* 32, 177-201.
- Levinson, Stephen 2016, Turn-taking in human communication - origins and implications for language processing. *Trends in Cognitive Sciences* 20, 1, 6-14.
- Reichle, Erik D., Reineberg, Andrew & Schooler, Jonathan W. 2010 Eye movements during mindless reading. *Psychological Science* 21, 9, 1300-1310.
- Rossano, Federico 2012 Gaze behavior in face-to-face interaction. Unpublished PhD, MPI for Psycholinguistics, Nijmegen.
- Rossano, Federico 2013 Gaze in conversation. In: J. Sidnell & T. Stivers (eds.), *The Handbook of Conversation Analysis*, Malden, MA: Wiley-Blackwell, 308-329.
- Rossano, Federico, Brown, Penelope, Levinson, Stephen C. 2009 Gaze, questioning and culture. In: J. Sidnell (ed.) *Conversation Analysis: Comparative Perspectives*. Cambridge: CUP, 187-249.
- Sacks, Harvey, Schegloff, Emanuel A. and Jefferson, Gail 1974, A simplest systematics for the organization of turn taking for conversation. *Language* 50(4), 696-735.
- Schegloff, Emanuel 2000, Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29, 1-63.
- Selting, Margret et al. 2011, A system for transcribing talk-in-interaction: GAT2 (translated and adapted for English by E. Couper-Kuhlen and D. Barth-Weingarten). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 12, 1-51 ([www. Gespraechsforschung-ozs.de](http://www.Gespraechsforschung-ozs.de)).
- Stivers, Tanja & Rossano, Federico 2010 Mobilizing response. *Research on Language and Social Interaction* 43, 1, 3-31.
- Stivers, Tanja, Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T. et al., 2009 Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. USA* 106: 10587-92.
- Streeck, Jürgen 2014 Mutual gaze and recognition: Revisiting Kendon's "Gaze direction in two-person conversation". In: Mandana Seyfeddinipur & Marianne Gullberg (eds.), *From Gesture in Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*. Amsterdam: Benjamins, 35-56.

- Stukenbrock, Anja 2018a, Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis. Mobile Eye Tracking-Analysen zur Herstellung von *joint attention*. Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion 19, 132-168. (www.Gespraechsforschung-ozs.de)
- Stukenbrock, Anja 2018b, Mobile dual eyetracking in face-to-face interaction: The case of deixis and joint attention. In: Geert Brône & Bert Oben (eds.) Eye-Tracking in Interaction. Studies on the Role of Eye Gaze in Dialogue, Amsterdam: Benjamins, 80-102.
- Stukenbrock, Anja forthc., Deixis, gaze practice and the interactive achievement of joint attention. To appear in: Frontiers in Psychology.
- Stukenbrock, Anja & Anh Nhi Dao 2019, Joint attention in passing: What dual mobile eye tracking reveals about gaze in coordinating embodied activities at a market. In: Elisabeth Reber & Cornelia Gerhardt (eds.), Embodied Activities in Face-to-Face and Mediated Settings. London etc.: Palgrave Macmillan, 177-216.
- Tiitinen, Sanni & Johanna Ruusovuori 2012 Engaging parents through gaze: Speaker selection in three-party interactions in maternity clinics. Patient Education and Counseling 89, 1, 38-43.
- Vranjes, Jelena 2018 On the role of gaze in the organization of turn-taking and sequence organization in interpreter-mediated dialogue, Language and Dialogue 8, 3, 439-467.
- Weiß, Clarissa 2018 When gaze-selected next speakers do not take the turn. In: Journal of Pragmatics 133, 28-44.
- Weiß, Clarissa 2019a Blickverhalten des nicht-blickselegierten Sprechers während Korrekturen und Elaborierungen. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion 20, 1-28. (www.Gespraechsforschung-ozs.de)
- Weiß, Clarissa 2019b Blick und Turn-Taking in Face-to-Face-Interaktionen. Multimodale Interaktionsanalysen triadischer Gesprächssituationen mit Hilfe von Eye-Tracking. Unpublished Ph.D., University of Freiburg.
- Zima, Elisabeth, Clarissa Weiß & Geert Brône 2019 Gaze and overlap resolution in triadic interactions. Journal of Pragmatics 140, 49-69.