Gestures Indicating Dialogue Structure

Hannes Rieser
Bielefeld University
`hannes.rieser@uni-bielefeld.de`

Abstract

This paper reports on a study carried out on the Bielefeld speech and gesture alignment corpus (SAGA). The study focussed on the problem, which of the gestures used by conversational participants (CPs) in the route and landmark-description dialogues support dialogue structure and what their specific function might be. Given that traditional gesture research mainly considered types of single gesture occurrences such as pointings or iconics, this is an entirely new perspective on the gesture-dialogue interface. On the dialogue description side the original conversation analysis (CA) account is used as a heuristics. Some discourse structure gestures found will be briefly commented upon. This will be followed by a description of four types of gestures: gestures used, respectively, to allocate next turn, in acknowledgements, for interrupts and a go-ahead gesture tolerating interrupts. It is argued that only integration of dialogue gestures brings dialogue theory down to real dialogue. Finally, it will be discussed how some aspects of the dialogue gestures can be integrated into recent versions of the Poesio-Traum-Theory of Dialogue (PTT).

1 Traditional Accounts of Gesture on the Gesture-Discourse Interface

Since its very beginning, gesture research was closely tied to natural discourse and dialogue but there is no research tradition linking empirically grounded formal description of gesture to theory of dialogue. Let us have a look on some of the leading scholars' work in the gesture and discourse field. McNeill (1992) focuses on narratives assuming that these are organised according to narrative, meta-narrative and para-narrative levels. 'Narrative' covers the main plot or story line, 'meta-narrative' categorizations of the structure of the narrative and 'para-narrative' relates to the observer's experience when confronted with the events narrated. In the short passage on conversations McNeill considers pointing with respect to topical information (pp. 216-217), which in terms of his levels belongs to the narrative one. In contrast to McNeill, Kendon (2004) has many examples of how gestures are used in different interactional moves on a local level but he does not group the units considered into larger structures; this may explain why the discourse function of gestures remained uncommented upon. The research closest to the interests pursued in the present paper is carried out by Bavelas and co-workers (1992, 1995). They provide experimental evidence for the existence of a subclass of conversational hand gestures, called interactive gestures, 'whose function is to aid the maintenance of conversation as a social system' (Bavelas *et al.* 1992, p. 470). These interactive gestures are used in the context of citing other's contributions, seeking help, marking information as new or shared or around turn organization (see the list in Bavelas et al. 1995, p. 397) Our research differs from the one of Bavelas et al. inasmuch as it focuses largely on the mechanisms of turn distribution proposed by classical CA and implemented in current versions of dialogue theory. We also differ with respect to methodology and the data considered: Our work is based

on the annotated and rated Bielefeld speech and gesture alignment corpus (SAGA), hence gesture meaning and function can be ultimately grounded in descriptions of fine-grained gesture morphology (cf. Lücking et al. 2010). Investigating the discourse function of gestures, we isolated a set of 1000 gestures out of SAGA's total 6000, to which two annotators ascribed discourse relevance (Hahn and Rieser 2009-2011). Differences notwithstanding, we sometimes have arrived at similar observations as McNeill, Kendon and Bavelas et al. In section two levels of situatedness of gestures in SAGA dialogues will be described. We give a short recap of turn-constructional rules in classical CA in section three. Section four describes some findings concerning gestures relevant for dialogue structure. Examples of those are provided in section five, turn allocation, acknowledgements, interrupts, and go-ahead!. In section six we look into quantitative data resulting from selected SAGA video films. In section seven we cast a PTT perspective on the corpus findings, to be followed by a short outlook on virtual reality (VR) simulation.

2 The SAGA Perspective: Situated Gesture

Fig. 1 shows a paradigm of the experimental scene on which the SAGA corpus is based. A Route-Giver[1] tells a Follower about a (VR-simulated) tour on a tourist bus through a town where he has followed a pre-fixed route passing five land-marks.

We have a face-to-face situation; the Follower may ask any question she likes. Now the Route-Giver may in principle use at least two information levels in his description: he may detail the route experienced (which has been VR) or he may use information from the situation in which the two CPs are in (which is real, being the so-called "cave", a device to record speech, CP's behaviour, eye-tracking data and the Route-Giver's gestures in R^4 space). Detailing the route implies for example describing the starting point, the route to the next landmark and so on, see Fig. 1 (b). The route is explained using the personal gesture space as a display: representations of routes and objects are placed into the gesture space. In contrast, using information from the situation means first of all exploiting one's and other's body and sometimes making use of various objects present in the immediate or larger situation such as the building where the experiment takes place or even the town Bielefeld. So, we have *analoga* to McNeill's narrative or Bavelas *et al.*'s topical information, but in addition, and that is the crucial point, there is the situation information exploited by the CPs. So, situatedness of gesture emerges with respect to two *loci*, the embedded gesture space and the larger embedding situation.

3 The CA Account of Turn Allocation

The early CA account of turn allocation (Sacks, Schegloff, Jefferson (1974)) although considered normative, is popular among theoreticians of dialogue (see for example Ginzburg 2011, to appear). It gives us the possibility to treat natural data, to extend CA findings to multi-modal dialogue and to preserve an interface to dialogue theory. First we provide the central turn-allocation mechanism in the original wording (p. 704): '3.3 RULES. The following seems to be a basic set of rules governing turn construction, providing for the allocation of a next turn to one party, and coordinating transfer so as to minimize gap and overlap.

1. For any turn, at the initial relevance place of an initial turn-constructional unit:

   a If the turn-so-far is so constructed as to involve the use of a 'current speaker selects next' technique, then the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place.

   b If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then self-selection for next speakership may, but need not, be instituted; first starter acquires rights to a turn, and transfer occurs at that place.

   c If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then current speaker may, but need not continue, unless another self-selects.

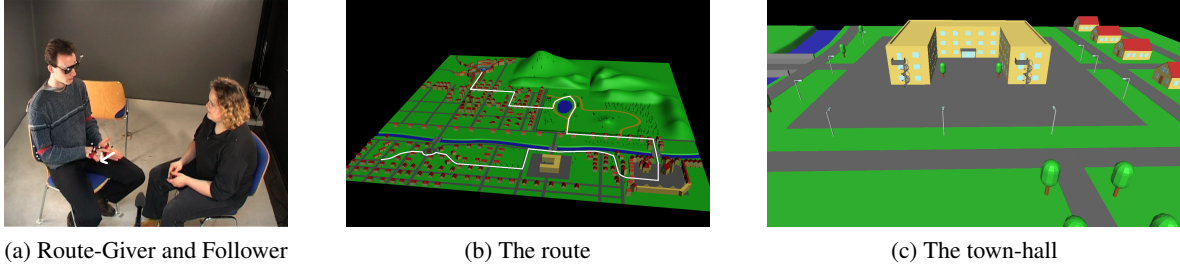| (a) Route-Giver and Follower | (b) The route | (c) The town-hall |

Figure 1: Experimental scene of the SAGA corpus

2. If, at the initial transition-relevance place of an initial turn-constructional unit, neither 1a nor 1b has operated, and, following the provision of 1c, current speaker has continued, then the rule-set *a-c* re-applies at the next transition relevance place, and recursively at each next transition-relevance place, until transfer is effected.'

We are well aware of Levinson's critique of these regulative mechanisms (see his 1983, pp. 294-371) but neglect it here as it would merit a paper on its own. Anyway, one should accept the fact that Sacks *et al.* argued that the turn distribution mechanism proposed is only a local device within an embedding speech exchange system. For example, in a prototypical SAGA route description dialogue the speech exchange system consists of the Route-Givers requests to the Follower concerning the Route-Giver's plan of the route and the descriptions of the land-marks which both have to be considered as perspective-oriented and plan-based. In addition, the speech exchange system also contains the systematic checks of the Follower who wants to be sure about the route taken and the landmarks encountered. So we have a mixture of dominating plan-oriented requests (Route-Giver), clarifications (Follower), repetitions (Route-Giver, Follower), revisions (Route-Giver, Follower) and acknowledgements (Route-Giver, Follower). The structure behind it is still ill understood but quite characteristic of much of natural task-oriented dialogue. Most probably, task-oriented dialogue is a mixture of different types of smaller speech-exchange systems, some of which might influence what is locally acceptable from the interactive point of view, see the remarks on overriding below. Anyway, concerning gesture, at least the following questions arise, given

the CA schema:

1. Can "current speaker selects next" be supported by gesture?
2. Can self-selection of next speaker be accompanied by gestures?
3. Is there a role for gesture to play in next turn?
4. How is non-orderly behaviour in dialogue treated gesturally?

Non-orderly behaviour might consist in in-turn short clarification requests or self-selection of next speaker "out of the normative order", i.e. under neglect of the preference as fixed in the CA schema. Which gestural markings might there exist on the side of current speaker or on the side of the intruding speaker?

Having prepared the ground for answers, we will comment on these questions again in section five. Actually, we discovered a lot of other things, as will be clear from the findings given in the next section but we will mainly concentrate on these questions in this paper.

4 Selected Findings

All of the gestures mentioned in the sequel are affiliated to speech, exceptions are indicated. Given the double situatedness of gestures explained in sect. 2, it is small wonder that indexing is used in SAGA in order to achieve turn allocation. Indeed, Bavelas *et al.* observed the social and communicative function of hand-shape (1992, 1995) before. As paradigm cases we have indexing of other to select other as next speaker. In addition, and this will be surprising, we encounter indexing of OTHER to select SELF as next speaker. Both will be documented in section five. So, gesture has an important role to play when it comes to determine who speaks next.

Viewed as grounding dialogue acts (Poesio and Traum, 1997), acknowledgements and accepts are of special relevance for pushing the dialogue forward: some sort of settledness must be achieved before the dialogue can go on. Acknowledging and accepting by other concerns the content of a previous dialogue act, so, the gestures used can be expected to have signifying power. Indeed, we find iconic gestures in second turns: the Follower imitates the Route-Giver's gesture or *vice versa*, if Route-Giver is second. Again, an example will be shown in the next section. There are still more types of gestural acknowledgements which we disregard here. However, a different matter is of interest: obviously, the discourse function of the iconic gestures is tied to their structural position in second turn; there are no *sui generis* iconic gestures for acknowledgements, in opposition to those indicating vagueness or low confidence in the information available. We will take up the question of *sui generis* gestures, which then might perhaps be viewed as emblems, shortly below.

In face-to-face construction dialogues correctness of the construction result must be tested. Similarly, route descriptions are characterized by short exchanges of information, often for purposes of control, checking a direction, the colour of a façade, the time shown on a public clock and so on. This is overriding the preferred CA order of interactive procedures. SAGA has at least two types of interrupts. There are those occurring in current speaker's mid-turn violating the rule that next turn is the privileged place for repairs or requests. In addition, we have cases of self-selection contravening the current speaker selects next rule.

Incidentally, the SAGA data show an important point as regards interaction in dialogue: violations do not go unnoticed and have to receive a treatment in terms of smooth interaction, un-orderliness thus being put into order. In more detail: First of all, current speaker can use "a don't interrupt gesture" if an impeding thrust of other is likely to come or has even just begun, for example, if current speaker hesitates but still wants to complete his turn. In contrast, a quick "out of order" interruption by other can be indicated by the other's "let me interrupt gesture" which is either a kind of pointing using G-shape or a slanted palm up directed against current speaker.

Finally, the currently interrupted speaker may react with a "go ahead!" response which is a kind of offering gesture with open palms cupped and upwards oriented towards the intruder. A complete interactional sequence of don't interrupt, let me interrupt and go ahead will be shown in the next section. In addition, we have interactive gestures like calming down and gestures indicating the truth-worthiness or the relevance of information from the Route-Giver's or the Follower's perspective but comments on those have to wait for another paper.

5 Substantiating Selected Findings: Four Types of Gestures Relevant for Dialogue, Turn Allocation, Acknowledgements, Interrupts and "Go-ahead!"

The extracts from the SAGA corpus presented by different Fig.s in this section give the following information: an excerpt from the multi-modal dialogue with CPs' German contributions, a translation into idiomatic English and the type of gesture used by RouteG or Follower. For example, Fig. 2 has a Follower gesture IndexingOthertoSelectSelf , Fig. 3 shows a normal (base-line, topical) pointing to the left of the Follower and Fig. 4 has an IndexingOthertoSelectOther. Annotation of gestures is according to SAGA standards (see Lücking et al. 2010 on that). The excerpt from the dialogue annotation also contains the information for the interface in which the gesture speech-integration[2] is defined. It is based on a time line not represented here. The role of the interface is easiest to explain with respect to the Follower's pointing to the left in Fig. 2: We have the words "to the left" which would receive a syntax representation in LTAG and a compositional semantics using $\lambda\beta$-calculus. This is fused with the gesture meaning also encoded in $\lambda\beta$. In the present case, gesture and speech have the same meaning and one of them is weeded out in the end. The typed attribute value matrices (AVMs) accompanying the stills show the respective stroke positions and specify the gesture morphological values.

All explanations given are based on SAGA annotations, which cause limitations of the following sort: Even if one could argue for underspecifi-

---

[2]The note on the interface tries to answer a question raised by reviewer 3.

cation and multifunctionality of gestures observed[3] this is not done in this paper because SAGA does not have systematic underspecification annotations. Anyway, we do not know of any strictly annotated multi-modal corpus dealing with underspecification in a systematic fashion. A problem not dealt with in this paper is how the AVMs are interpreted semantically. In short, they are mapped onto a partial ontology. This way, iconic gestures receive their own (Peircian, if you like) meanings. Gesture meaning is compositionally fused with verbal meaning, in this way gesture-speech ensembles get a unified meaning (see Rieser 2010 and 2011a on that and Giorgolo 2010 and Lücking 2011 for different options).

5.1 Indexing Other to Select Other

Speech & gesture
Route-Giver: Das nächste Ziel das ist ja   dann
RG.: The next   goal that is well then
RG.-Gesture:
RG.: quasi   die Endstation, Richtung
RG.: effectively the final stop, direction
RG.-G.: `Turnkeep-`
RG.: Brunnen.   **Follower:** An der Kapelle
RG.: fountain.   **F.:** At the chapel
RG.-G.: `ing`   **F.-Gesture.:** `IndexOther-`
F.: geht's jetzt aber links ab
F.: it now branches off to the left
F.-G.: `ToSelectOther   Indexing`
F.: nicht   rechts ab
F.: though not to the right.
F.-G.: `IndexingToSelectOther`

V5 11.50 *IndexingSelect*	
HandShapeRH	G
BOHDirectionRH	BAB
PalmDirectionRH	PTL
WristPositionRH	CUR
WristPosDistRH	DEK
TargetRH	CC-other

Figure 2: Follower's IndexingOtherToSelectOther

V5 11.51 *Index*	
HandShapeRH	G
BOHDirectionRH	BTL
PalmDirectionRH	PTB
WristPositionRH	CUR
WristPosDistRH	DEK

Figure 3: Follower's Topical Indexing of Left

---

[3]Reviewer 2 made the point on underspecification and multifunctionality referring to work by McNeill, Kendon and Bunt.

V5 11.54 *IndexingSelect*	
HandShapeRH	G
BOHDirectionRH	BAB
PalmDirectionRH	PTL
WristPositionRH	CUR
WristPosDistRH	DEK
TargetRH	CC-other

Figure 4: Follower's Second IndexingOtherToSelectOther

The Route-Giver reports about approaching the final landmark, a fountain. He indicates using a "thinking about" emblem that he wants to keep the turn. The Follower pointing at the Route-Giver is going back to a previous stage in the route traversed. The matrix shows the familiar gesture morphology of indexing except its orientation to CC-other. This is crucial. It maps pointing into the larger situation (see section 2). We have a clarification request of the Follower. Why is the Follower entitled to issue a clarification request? Besides, it is a self-selection by way of an interrupt, especially in view of the turn-keeping gesture of the Route-Giver. Could there be another interpretation of the Follower's action or an additional one? A reprimand of some sort perhaps? This would not change much, only add an interactional component. Anyway, we clearly see the difference between pointing at other (Follower) and indexing a direction to the left in the Follower's gesture space (Figure 3, no CC-other target). Note that in pointing to the left, the Follower's perspective is "myself on the route". However, there is a shift after the pointing to the left to the indexing of other again. So, we also have a shift from the inner to the outer situation.

5.2 Indexing Other to Select Self

Speech & gesture
RG.: Links   steht ne Kirche, rechts
RG.: To the left is   a church to the right
RG.-Gest: `Indexing        Indexing`
RG.: steht ne Kirche. **Follower:** Moment
RG.: is   a church. **Follower:** Wait
RG.-Gest. `Indexing cnt'd` **F.-Gest.** `IndSlctSlf`
F.: noch eine Frage zum   Rathaus.
F.: a question   as regards the townhall
F.-G.: `Indexing`
F.: Welche Farbe?
F.: Which color?
F.-G.: `Indexing cnt'd`

13

$$
\begin{bmatrix}
\textit{V8 3.20 IndexOthToSelSelf} \\
\text{HandShapeRH} \quad \textit{H} \\
\text{BOHDirectionRH} \quad \textit{BAB/BUP} \\
\text{PalmDirectionRH} \quad \textit{PTL} \\
\text{WristPositionRH} \quad \textit{C-RT} \\
\text{WristPosDistRH} \quad \textit{DEK} \\
\text{TargetRH} \quad \textit{CC-other}
\end{bmatrix}
$$

Figure 5: Follower's IndexingOthertoSelectSelf

$$
\begin{bmatrix}
\textit{V8 3.22 Indexing} \\
\text{HandShapeRH} \quad \textit{tapered O} \\
\text{BOHDirectionRH} \quad \textit{BAB/BUP} \\
\text{PalmDirectionRH} \quad \textit{PAB/PDN} \\
\text{WristPositionRH} \quad \textit{P-RT} \\
\text{WristPosDistRH} \quad \textit{DEK}
\end{bmatrix}
$$

Figure 6: Follower's non-canonical Indexing

The Route-Giver has already reached the churches and indexes both. Follower self-selects (Fig. 5, a milder transgression than in the first case), accompanying it with indexing other. The matrix for the gesture shows that the Follower's back of right hand is slanting up and we have the other as target (CC-other). There is a similarity to the first example inasmuch as the Follower is backtracking on the route to the last landmark (see town-hall, Fig. (1c)). Again, selecting next speaker is followed by a sort of non-canonical indexing (Fig. 6).

5.3 Acknowledgement

Speech & gesture
RG.: [Das heißt] es hat vorne   so
RG.: [That is]   it has to the front kind of
RG.: zwei Buchtungen und   geht hinten
RG.: two bulges   and closes in the rear
RG.-Gest. Shaping
RG.: dann.
RG.: then.
RG.-G. cnt'd
F.:   OK.
F.:   OK.
F.-G.: nod

Figure 7 matrices:

V8 1.50 Shaping		*V8 1.50 Acknowledgement*	
HandShapeRH	*small C*	HandShapeRH	*loose C*
BOHDirctRH	*BAB/BTR> BAB>BAB/BTL*	BOHDirectRH	*BAB> BAB/BTR*
PalmDirctRH	*PAB/PTL> PTL>PTB/PTL*	PalmDirectRH	*PTL> PAB/PTL*
WristPosRH	*CR*	WristPosRH	*CC>CR*
WristPosDisRH	*DEK*	WristPosDisRH	*DEK*
HandShapeLH	*small C*	HandShapeLH	*loose C*
BOHDirctLH	*BAB/BTL*	BOHDirectLH	*BAB> BAB/BTL*
PalmDirctLH	*PAB*	PalmDirectLH	*PTR> PAB/PTR*
WristPosLH	*CL*	WristPositionLH	*CC>CL*
WristPosDistLH	*DEK*	WristPosDistLH	*DEK*
WristMovRH	*MF>ML*	WristMovRH	*MR>MB*
WristMovLH	*MF*	WristMovLH	*ML>MB*
PathOfWristRH	*Line>Line*	PathOfWristRH	*ARC*
PathOfWristLH	*Line*	PathOfWristLH	*ARC*
TwoHandMove	*mir-sagit>∅*	TwoHandedMove	*mir-sagit*

Figure 7: Route-Giver shapes the town-hall (left AVM) and Follower imitates gesture (right AVM)

The Route-Giver describes the shape of the town-hall (left AVM). The Follower acknowledges taking up the iconic gesture of the Route-Giver (right AVM). Observe that the gestures are different, since both agents gesture the parts of the town-hall in a different way. This can be seen from the information in the respective matrices comparing them line by line. However, if the respective gesture-parts used by the Route-Giver and the Follower are compared, both gestures yield the same set of three-block buildings and both are satisfied by the VR edifice (see Fig. (1c)). Note that the Follower changes the perspective, signing as if she were already at the place looking into the court of the town-hall. What she produces is an implicit anaphora to the Route-Giver's *it* and a copy of the property as gestured by the Route-Giver.

5.4 Don't Interrupt, Let me Interrupt, Go ahead

Speech & gesture
RG.: Gerade aus, gut. Moment.
RG.: Straight on, OK. Just a moment.
RG.-Gest: DontInterrupt
RG.-G.: Don'tInterrupt cnt'd
F.: Ich vollzieh nochmal den Weg nach.
F.: I'll recapitulate once more the route.
F.-G.: LetMeInterrupt
RG.: Ja.
RG.: Ok.
RG.-G.: Go Ahead

V4 2.19 DontInter

HandShapeRH	*loose B spread*
BOHDirectionRH	*BAB/BUP*
PalmDirectionRH	*PTL*
WristPositionRH	*PLW*
WristPosDistRH	*DEK*
HandShapeLH	*loose B spread*
BOHDirectLH	*BAB/BUP*
PalmDirectLH	*PTR*
WristPositionLH	*PLW*
WristPosDistLH	*DEK*

V4 2.21 LetMeInter

HandShapeRH	*loose B spread*
BOHDirectRH	*BUP*
PalmDirectRH	*PTL*
WristPositionRH	*C-RT*
WristPosDistRH	*D-CE*
HandShapeLH	*loose B spread*
BOHDirectLH	*BUP*
PalmDirectLH	*PTR*
WristPositionLH	*CC*
WristPosDistLH	*DEK*
TwoHandConfig	*PF*

V4 2.23 GoAhead

HandShapeRH	*loose B spread*
BOHDirectionRH	*BAB*
PalmDirectionRH	*PDN/PTL*
WristPositionRH	*CLR*
WristPosDistRH	*DEK*
HandShapeLH	*loose B spread*
BOHDirectionLH	*BAB*
PalmDirectionLH	*PDN/PTR*
WristPositionLH	*CLL*
WristPosDistLH	*DEK*
TwoHandConfig	*PF*
TargetRH	*CC-other*
TargetLH	*CC-other*

Figure 8: Sequence of Route-Giver's DontInterrupt (1st AVM), Follower's LetMeInterrupt (2nd AVM), and Route-Giver's GoAhead! (3rd AVM)

Here the Route-Giver describes a direction. Then she stops indicating that she has to think about the next step (*Just a moment*). She produces a sort of warding off gesture barring off the Follower's intrusion (1st AVM). The Follower issues a "let me interrupt gesture", some thrust forward indicating already the direction to go, and starts his interruption (2nd AVM). He suggests a recap of the route. The warding off gesture of the Route-Giver continues during this indication. Then we have an accept of the Route-Giver, OK, and a "go ahead gesture" indicated by a sort of handing over turn-production to the Follower (3rd AVM).

5.5 Summary of Results

Finally, we take up the questions posed in section three and provide answers substantiated by the data we have seen: Current speaker's selecting next can be supported by his pointing gesture, as can self-selection as next speaker. Gesture in next turn may be used for purposes of acknowledgement, more in general, for purposes of providing feed-back. Non-orderly behaviour can be fenced off by current speaker. If producer of non-orderly behaviour insists, he may be granted execution. So the non-orderliness problem is smoothened out on the level of interaction.

6 Some Figures

Fig. 9 presents a few figures divided up into those valid for the whole SAGA corpus (above) and those we get from the re-annotated data V1 – V15 (below). In the total corpus we have 7437 gestures including moves. Moves are dynamic gesticulations we could not classify at some point of the on-going annotation procedure. These are considered prime candidates for re-annotation. In the corpus there were 3165 iconic, 1311 deicitc, 1223 discourse and interaction gestures and 929 mixed occurrences overlapping gesture types or practices, for example the types iconic and deictic or the practices indexing and shaping. Observe the high number of discourse gestures in the original data. In the re-annotated material there are 2570 gestures including moves, candidates for re-annotation, and 320 discourse gestures. Of the discourse gesture types discussed in sections four and five, there were 12 occurrences of IndexingOthertoSelectOther, 11 of IndexingOthertoSelect-

Figure 9: Number of Gestures in SAGA

Corpus (25 video films)					
Gestures incl. Moves	Iconic	Deictic	Discourse Gestures	Mixed	
7437	3165	1311	1223	929	
Videofilms Anotated for Discourse & Interaction Gestures					
Gestures incl. Moves	Iconic	Deictic	Discourse Gestures	Mixed	
2570	1301	377	320	309	
Indexing Other to Select Other	Indexing Other to Select Self	Acknowledgement w. Imitation	Don't Interrupt	Let me Interrupt	Go ahead!
12	11	21	1	6	5

Self, 21 of AcknowledgementwithImitation, 1 for Don'tInterrupt, 6 for LetmeInterrupt and 5 for Go ahead!. The rest, adding up to 320 gestures, was not used for this paper.

7 A PTT Perspective on the SAGA Corpus Findings

The gesture data discussed in section 5 put different demands on a theory like PTT. In PTT we have already dealt with the modelling of pointing (Rieser and Poesio 2009), anaphora (Poesio and Rieser 2011), completions and repairs (Poesio and Rieser 2010) but not with the kind of layered situation information as described in section 2. We have to consider the distinction made in section 2 allowing for two kinds of situatedness, one being given by the outer situation (cave and surroundings) and the other as displayed in the gesture spaces proper. If the agent is the target of a gesture, then it is the outer situation that matters, if, in contrast, we remain on the level of the discourse information, the gesture space is prevalent. Deferred reference is a good rule of thumb in this respect. The shift from inner to outer situation and vice versa (we'll see an example of that from SAGA, V5 11.52 below) is indicated by targeting the body of the addressee or a location in the gesture space. However, there is a commonality between both types of cases: we must always take account of the visual situation, in other words, the visual situation is the resource situation of meaning expressed gesturally. We take SAGA, V5 11.52 (see 5.1, IndexingOthertoSelectOther) as an example in order to show the problems arising. Here is a list of phenomena we have to deal with on the Route-Giver's and the Follower's side:

(1a) Route-Giver's anaphora *the next goal* and

(1b) his turn-keeping gesture "Let me think";

(2a) Follower's selecting Route-Giver as next speaker by indexing,

(2b) Follower's production of a clarification request and

(2c) his use of an anaphor *it*,

(2d) Follower's use of topical indexing while recapitulating part of the route, thus dealing with the inner situation.

(2e) Follower's switch back from the inner to the outer situation to select Route-Giver again as next speaker by pointing.

We will use a simplified version of PTT here and avoid discussing problems of precise timing of speech-gesture-interfaces, micro-conversational events, anaphora, and incrementality (see Rieser and Poesio 2009, Poesio and Rieser 2010 and 2011 for

more detailed explanations and formalisations). Relevant sections of the example will be indexed with (1a) - (2e) introduced above.

```
       DU1 is
       [[K1.1, ce1.1, up1.1|
        up1.1: utter(Route-Giver,"The next goal is ⟦ well then effec-
        tively ⟧
                            the final stop"),
        sem(up1.1) is K1.1,
        ce1.1: assert (Route-Giver, Follower, K1.1),
  (1a)  K1.1 is [g, f, s|g is [ɿxK;[|goal(x), next(x)]];
                        s is [ɿyK;[| final(y), stop(y), direction-of(y,f),
                                            fountain(f)]],
                        s is g],
        generate(up1.1, ce1.1)],
       [K1.2,Kmva1.2, ce1.2, gp1.2|
        K1.2 is Kmva1.2,
  (1b)  gp1.2: gesture(Route-Giver, "Let-me-think"-emblem)
        sem(gp1.2) is K1.2, ce1.2: request(Route-Giver,Follower,K1.2),
        K1.2 is [| e: let-go-on(Follower, Route-Giver)],
        generate(gp1.2, ce1.2),
        overlap(up1.1, gp1.2)]]
  (2a)  DU2 is
       [[K2.1, Kmva2.1, K2.2, K_{dt}, K2.3, Kmva2.3, ce2.1,
        pe2.1, up2.2, ce2.2, ce2.3, pe2.3|
        pe2.1: point-at(Follower,Route-Giver),
        sem(pe2.1) is K2.1, ce2.1: indicate(Follower, Route-Giver,
        K2.1),
        K2.1 is Kmva2.1,
        Kmva2.1 is [|e: select-as-next-speaker(Follower,Route-Giver)],
        generate(pe2.1, ce2.1),
  (2b)  up2.2: utter(Follower, "At the chapel it now branches )
                            off to the left, ⟦though⟧,
                            not to the right"),
        sem(up2.1) is K2.2,
        ce2.2: quest(Follower, Route-Giver, K2.2),
  (2c)  K2.2 is [x, l, r, pe2.2|
        x is [ɿy K_{dt}; [|route(y)]],
        e: branch-off(x, l), to-left(l),
        e: ¬branch-off(x, r), to-right(r)],
        p2.2: point(Route-Giver, l),
        generate(up2.2, ce2.2),
  (2d)  pe2.3: point(Follower,Route-Giver),
        K2.3 is Kmva2.3,
        sem(pe2.3) is Kmva2.3,
        ce2.3: indicate(Follower, Route-Giver, K2.3)
  (2e)  K2.3 is [|e: select-as-next-speaker(Follower,Route-Giver)],
        generate (pe2.3, ce2.3),
        overlap(pe2.1, "At the chapel it branches off"),
        overlap(pe2.2, "to the left"),
        overlap(pe2.3, "⟦though⟧ not to the right")]]
```

Following Potts (2005), we consider expressions in "⟦,⟧" to be treated as Conventional Implicatures and the rest as "at issue" material.

(1a): *next* is a presupposition trigger, the resource situation K yielding the presuppositional information and similar for *final*.

(1b): To capture the "Let-me-think" gestural emblem expressing a request, we need a situation of mutual visual attention, *Kmva1.2*.
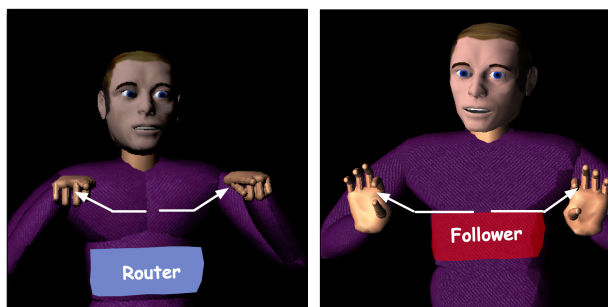
(2a): Being the current speaker, the Follower points at the Route-Giver to select him as next speaker.

(2b): Part of the Follower's clarification request *up2.2*, "at the chapel it now branches off to the left?", overlaps with his pointing *pe 2.1* at the Route-Giver.

(2c): In contrast, his clarification request contains a pointing *pe 2.2* which is topical information and overlaps a different part of the clarification, namely "to the left". The difference between inner and outer situation is reflected in the difference of the arguments to the pointing, Route-Giver *versus* locational direction argument *l*. Resolution of the anaphor *it* demands that we introduce a resource situation K_{dt} providing us with the property *route* and a discourse referent *x* having this property.

(2d, 2e): The Follower again selects Route-Giver as next speaker by pointing, supporting his addressed clarification request.

8 Evaluating Gesture Dialogue Interfaces Using VR Simulation

The example in sect. 6 is based on studies dealing with SAGA, hence, a number of intermediate stages led from initial annotation (see sect. 5) to final encoding in PTT. How can one falsify a product resting on so many preconditions? The answer lies in simulation by VR avatars (see Fig. 10). We already started to simulate the acknowledgement in 5.3 based on the material discussed there (Bergmann, Rieser, and Kopp 2011). The focus of the simulation is speech-gesture timing and avatar's gestures mimicking the gesture morphology used by the CPs. This provides us with a fairly precise measure of the correctness of the research line taken and indicates where things have to be improved at the various stages.

Figure 10: Simulation of Router and Follower gestures from *5.3 Acknowledgement* using VR avatar Max (cf. Bergmann, Rieser, Kopp 2011).

References

Bavelas, J., Chovil N., Lawry, D., Wade, A. (1992). Interactive gestures. In: Discourse Processes 15, pp. 469-489

Bavelas, J., Chovil N., Coated, L., Roe, L. (1995). Gestures Specialised for Dialogue. In: PSPB, vol. 21 No. 4, pp. 394-405

Bergmann, K., Rieser, H. and Kopp, St. (2011). Regulating Dialogue with Gestures - Towards an Empirically Grounded Simulation with Conversational Agents. SIGdial 2011

Ginzburg, J. (2011). The Interactive Stance. Meaning for Conversation. OUP (in print)

Giorgolo, G. (2010). Space and Time in Our Hands. LOT, Utrecht Institute of Linguistics.

Hahn, F. and Rieser, H. (2009-2011): Dialogue Structure Gestures and Interactive Gestures. Annotation Manual. CRC 673, Alignment in Communication. Working Paper, Bielefeld University

Kendon, A. (2004). Gesture. Visible Action as Utterance. Cambridge: CUP

Levinson, St. C. (1983). Pragmatics. CUP

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010). The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In M. Kipp et al. (Eds.), LREC 2010 Workshop: Multimodal Corpora-Advances in Capturing, Coding and Analyzing Multimodality, pp. 92-98

Lücking, A. (2011). Prolegomena zu einer Theorie ikonischer Gesten. Phil. Diss. Bielefeld University.

McNeill, D. (1992). Hand and Mind. ChUP

Poesio, M. and Traum, D. (1997). Conversational Actions and Discourse Situations. In: Computational Intelligence 13 (3), pp. 309-347

Poesio, M. and Rieser, H. (2009). Anaphora and direct reference: Empirical evidence from pointing. In: Proceedings of DiaHolmia, the 13th Workshop on the Semantics and Pragmatics of Dialogue, Stockholm,

Poesio, M. and Rieser, H. (2010). Completions, coordination, and alignment in dialogue. Dialogue and Discourse, 1(1), 1-89

Poesio, M. and Rieser, H. (2011). An Incremental Model of Anaphora and Reference Resolution Based on Resource Situations. Dialogue and Discourse, 2(1), 235-277 Potts, Chr. (2005). The Logic of Conventional Implicatures. OUP

Rieser, H. (2010). On Factoring out a Gesture Typology from the Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). In S. Kopp and I. Wachsmuth (Eds.), Proceedings of GW 2009: Gesture in Embodied Communication and Human-Computer Interaction (LNAI 5934, pp. 47-60). Berlin/Heidelberg: Springer.

Rieser, H. (2011). How to Disagree on a Church-Window's Shape Using Gesture. In: Hölker, K., Marello, C. (Eds.), Dimensionen der Analyse von Texten und Diskursen. Festschrift für Janos Sandor Petöfi zum achzigsten Geburtstag. Münster: LIT Verlag, pp. 231-247

Rieser, H., Poesio, M. (2009). Interactive Gesture in Dialogue: a PTT Model. In: Healey, P. et al. (eds), Proceedings of the SIGDIAL 2009. London, ACL, pp. 87-96.

Sacks, H., Schegloff, E., Jefferson, G. (1974): A simplest systematics for the organization of turn-taking for conversation. In: Language Vol. 50, pp. 696-735