Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology

Florian Hahn and **Hannes Rieser***
CRC 673, *Alignment in Communication*, Bielefeld University

Abstract

The paper discusses how a gesture typology can be extracted form the Bielefeld Speech-And-Gesture-Alignment corpus (SAGA) making use of the annotated gesture morphology in the SAGA data. The SAGA corpus is briefly characterized. Using a portion of a MM dialogue, the interface between speech and gesture is shown focussing on the impact of gestures on lexical definitions. The interface demonstrates the need for working out types of gestures and specifying their semantics *via* a partial ontology. This is started setting up a "typological grid" for 400 gestures. It yields a hierarchy of *n*-dimensional single gestures and composites of them. A statistical analysis of the grid results is provided and evaluated with respect to the whole SAGA corpus. Finally, it is shown how the typological results are used in the specification of the speech-gesture interfaces for the MM dialogue.

Keywords: completion, gesture typology, gesture-speech interface, partial ontology, SAGA

1 Motivation, Dialogue Example from SAGA, Plan of Paper

Do referential and iconic gestures have a specifiable meaning and function in multi-modal dialogue? Questions like these are normally handled in a descriptive way with respect to arbitrarily collected empirical data, for example in the gesture research based on a semiotic tradition as initiated by Ekman and Friesen (1969) and carried on in McNeill (1992) and Kendon (2004). We have built up a corpus of multi-modal data, the Bielefeld Speech and Gesture Alignment corpus, SAGA (see Lücking et al. 2010), completely annotated for referential and iconic gestures to deal with them in a systematic way. Below we give a short characterization of SAGA. We investigate several topics with respect to SAGA, focussing on gestures co-occurring with noun-phrases in MM dialogue:

(a) the types of gestures used and their function, for example production of a one-dimensional line,

(b) the semantic value a gesture represents like being the boundary line of an object,

(c) how the semantic value interfaces with an accompanied natural language expression such as *the church-window* and,

(d) how a multi-modal meaning can be built up from the meaning of the words and the gesture's meaning in a compositional way,

(e) how MM meanings behave in dialogue.

We want to answer question (b), the semantic value represented, serving as a precondition for answering (c) to (e), hence the interest in systematic typology. At the outset, we present a dialogue example (Fig. 2) being about two churches and their windows (Fig. 1) which illustrates the function of gesture and which will serve as our reference datum throughout the paper.

The specific SAGA setting for the example used has the following caracteristics: Two

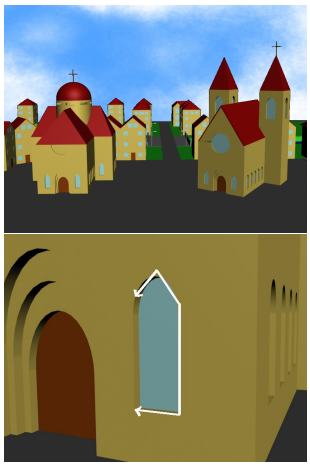

Figure 1: Looking at the VR-representation of the churches involved. The VR-stills show the two churches mentioned in the dialogue passage of Fig. 2. The Router's idea that the windows are gothic seems to come from the shadowing in the VR window representation. Anyway, the gesture used to depict the windows is fairly correct. This is indicated by the merge of the VR-window and the Router's gesture in the bottom still.

*Names in alphabetical order, 'MM' abbreviates 'multi-modal'. Corresponding author: Hannes Rieser, Hannes.Rieser@Uni-Bielefeld.de

R1 R2 R3

Dialogue Part 1

Router-Speech:	Beide	Kirchen	haben	diese	typischen	Kirchenfenster,	halt	unten
	Both	churches	have	these	typical	church-windows,	you know,	at the bottom
Router-Gesture:						Hedging	R1	R1
Follower-Speech:								
Follower-Gesture:								

Router-Speech:	eckig,	nach oben	dann	halt	so		gotisch	zulaufend.
	cornered,	upwards	moreover	you know	so		gothically	pointy
Router-Gesture:	R1	R2	R2	R2	R3			
Follower-Speech:						gotisch		
						gothic		
Follower-Gesture:			Nodding				Nodding	

Figure 2: The Router describing the church-windows to the Follower with the crucial bend at the top (R3).

agents, a Router, wearing body trackers, describing a route through a VR town and a Follower, trying to take up his description, face each other. The Router describes a landmark, two churches with gothic windows. The dialogue passage in Fig. 2 contains his multi-modal description of the church-windows. We present the German wording, the English translation and the stills showing the gestures in the way they accompany the words; we will rely on the English translation in the rest of the paper. The churches involved are depicted in Fig. 1.

What can we observe in the dialogue example? We treat the Router's contribution and the Router-Follower interaction.

The Router's Contribution: The Router starts gesturing parallel to wording *church-window*. Intuitively, he marks with his left hand (LH) the onset of an object and draws a line with his right hand (RH). The two-handed gesture R1 depicts some sort of corner. This goes on until the production of *upwards*. Concurrently with *upwards* his RH goes back to the onset held with LH and moves up, generating R2. Parallel to *so* the Router draws a hook R3, the top of a church-window.

The Router-Follower-Interaction: The Router's drawing is faster than his speech production, in addition, the German *halt* + *so*, engl. *you know* + *so* acts as a hesitation signal. Here the Follower comes in with her completion *gothic*. This is repaired by the Router extending it with *gothically pointy* (cf. Fig. 2).

So much for the data. We'll discuss the empirical findings from the point of view of corpus-based gesture typology and partial ontology. Section 2 provides background information concerning SAGA. Then we come back to the dialogue example again, investigating what we need for explaining speech gesture alignment in the dialogue (Section 3). Three statistics sections follow, one on the methodology of the typological grid (Section 4), the other on the statistics of it giving the frequency of types of gestures such as pointings, lines etc. in the grid (Section 5), and the third one an evaluation on how the statistical results for the grid can be generalized for the whole SAGA corpus (Section 6). Finally, we describe how the typology is used to provide explanations for the MM

dialogue part in Fig. 2 (Section 7).

2 Background on SAGA

The SAGA corpus contains 25 route-description dialogues taken from three camera perspectives. The setting comes with a driver, called "Router" "riding in a car" through a VR-setting, passing five landmarks connected by streets. After his ride the Router relates his experience in detail to a Follower supposed to organise her own trip following the Router's instructions. We collected video and audio data for both participants, for the Router in addition body movement tracking data due to markers on head, wrist, and elbow and eye-tracking data. The tracking data generated traces in Euclidean space and provided exact measurements for positions of head, elbow, and wrist. The gestures in the dialogues have all been annotated, the annotation predicates used like *indexing, modelling, shaping etc.* were rated.

The rating of these predicates is discussed in (Lücking et al. 2010). An example of a partial gesture annotation is given in Appendix, Fig. 3. The SAGA corpus contains ca. 6000[1] gestures. Roughly 400 gestures have been investigated in order to establish the typological grid described below.

3 What we Need for Explaining Speech-Gesture-Alignment in the Dialogue Example

3.1 The Router's Gestures

Consider Fig. 2. In R1, the Router's LH marks the left side of an object. Then he draws a horizontal line in three-dimensional space. R2 marks a vertical line starting at the corner set by the horizontal line and the left hand side producing a lower right angle. The vertical line ends in a fairly sharp bend. The only completely designed object is the bottom left "corner", all the other depictions are incomplete or underspecified.

3.2 Notes on the Speech-gesture Interface

We assume that gestures can be equipped with meaning, as a rule with a partial one and that gesture meaning interfaces with verbal meaning on different structural levels. Consequently, some gestures go with word meanings, others with the meaning of constituents, the meaning of dialogue acts, of rhetorical relations and so on. In previous papers we have shown how these various interfaces can be constructed (Rieser (2004, 2010), Rieser and Poesio (2009)) using type logics and compositional DRT. In this paper we will remain at the lexical and syntactic level. So, what we have to do is to observe the alignment of speech and gesture in these three cases all bound up with *church-window*:

Case 1:	you	know,	at	the	bottom	cornered
			R1	R1		R1
Case 2:	upwards,		moreover,		you know	
			R2		R2	R2
Case 3:	so					
	R3					

We consider the two occurrences of *you know* as meta-communicative acts[2] which can be inserted in multi-modal face-to-face communication before any complex constituent. *You know* can e.g. be taken as an attention-securing device, asking for acknowledgement or as focussing the relevant encyclopedic knowledge, whichever, it should be separated from the constituents contributing to the information describing the relevant event, i.e. in our case the Router's ride. So we have to care for the speech-gesture alignment of the rest.

3.3 Fine Grained Word Meaning, Partial Ontology, and Construction Meaning

After these considerations we turn to the details, first to the syntax of the router's contribution. The relevant part is shown in Appendix, Fig. 4. We have an N followed by two attribute-phrases AttrPhrs conjoined by *moreover*. The construction in Appendix, Fig. 4 is incomplete: It closes with an AdjPh where only the Adv *so* is realized. A completion is produced by the Follower. It is the Adj *gothic* which completes the AdjPh, see Fig. 4 for the cooperatively produced attribute. One should be aware of the fact that the scope of the completion is the "gappy" attr-phrase under construction by the Router. Roughly, we have to fuse in the end the meanings of *cornered at the bottom* and R1, *upwards* and R2, and *so* and R3. Below we provide lexicon definitions for *church-window*, and the morphology of the gestures R1, R2, R3 in-

[1] Due to filling gaps in our annotation and due also to re-annotation of material having already been annotated, both of which change the segmentation of gestures, the figures given for the number of gestures in SAGA have changed somewhat, roughly from 5000 plus to 6000.

[2] A paradigm which focuses on "inter-leavings" of meta-communicative and base-line material is Ginzburg (2010).

volved with their respective partial ontologies associated (*cf.* ch. 7).

Due to limits of space we concentrate in this paper on the lexicon definition of *church-window* and how it interfaces with the gestures R1-R3 without going into much detail.

Church-window: $\textit{church-window}(w) := \textit{window}(w) \wedge \textit{part-of}(w, wa) \wedge \textit{wall}(wa, ch) \wedge \textit{church}(ch) \wedge \textit{lower-part-of}(lp, w) \wedge \textit{middle-part-of}(mp, w) \wedge \textit{upper-part-of}(up, w) \wedge lp \oplus^{3} mp = cub \wedge \textit{cuboid}(cub) \wedge \textit{base}(b1, cub) \wedge \textit{breadth}(br, b1) \wedge \textit{breadth}(br, s1) \wedge \textit{side}(s1, lp) \wedge \textit{upper-part-of}(up, w) \Longleftrightarrow (((\textit{prism}(up) \wedge \textit{pointed}(up) \wedge \textit{acute}(up)) \vee (\textit{cylindric-section}(cls, up) \wedge \textit{round}(cls)))$.

A church-window *w* is part of a church *ch's* wall *wa* having a lower *lp*, a middle *mp* and an upper part *up*; the lower and the middle part form a cuboid *cub* with a base *b1* of a certain breadth *br* and a *side1*, part of *lp*, of the same breadth, the upper part *up* being either a pointed prism or a cylindric section. This gives us two versions of a church window, a gothic style one and a round arched one as depicted in Fig. 5 and Fig. 6, respectively.[4]

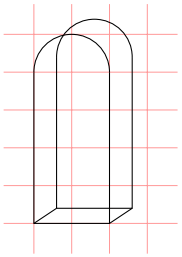

Fig. 5: The church-window as depicted by the right term of the disjunction $(\textit{cylindric-section}(cls, up) \wedge \textit{round}(cls))$

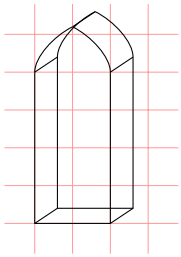

Fig. 6: The church-window as depicted by the left term of the disjunction $(\textit{prism}(up) \wedge \textit{pointed}(up) \wedge \textit{acute}(up))$

Fig. 7 represents the interface of speech and gesture in dialogue part 1. The arrows outside the pictures pointing towards the lexicon definitions indicate that gesture content operates on lexical content. In order to model that we would need the most basic gestural features and how they compose to make up gestures and their content. However, are there basic gesture features and complex ones observed by one speaker or even all speakers at all?

4 The Methodology of the Typological Grid

The question we ended up with in the previous section is: Are the gestures observed, sides, lines, the three-dimensional entities arising, arbitrary tokens, sporadic whims of the CPs' or are these systematically used at least throughout one datum (here video-film V5) by two agents or throughout the whole SAGA corpus by many or even all agents? In order to investigate both these typological questions we have set up a so-called typological grid for the datum V5 in the following way: intuitively, gestures build a space, consisting of simple and more complex gesture-morphological entities. The most fundamental entities we have are the individual annotation predicates like *hand-shape*, *wrist-movement* or *palm-direction*. For example, for the horizontal line starting at the left corner of the object in the example (see Fig. 2, R1 - R3, LH), we need the individual predicates *hand-shape*, *wrist-movement to the right* and *palm-direction facing down* or *left*. These are atomic features of the gesture space represented by AVM-matrices (App., Fig. 9). Only taken together do these single bits of information describe a horizontal line, again represented by a matrix (App., Fig. 10). The atomic features taken together form the most basic stratum of the gestural space (App., Fig. 17). They are derived from rated annotation predicates. Next in terms of complexity are the clusters: Clusters are bundles of functionally connected features, cf. section 5.2 for further motivation. 0-dimensional entities are introduced *via* the *indexing* annotation predicate. 1-dimensional entities originate from annotation predicates *drawing* or *modelling* and from wrist-movements. 2-dimensional entities, i.e. locations, regions and "more geometrical" 2D-objects are founded on *indexing*, *drawing* and *shaping*. 3-dimensional entities are based on *placing*, *shaping*, *drawing* or *modelling*.

Lines come with different bends and directions. They form the one-dimensional layer below the

[3] '$\oplus$' indicates mereological addition

[4] One of the reviewers is right in assuming that the lexicon definition is too near to the corpus data. A way out would be to leave the upper section of the church-window underspecified as to shape and to assume that it is the Router's gesture which resolves the underspecification with respect to *pointed* and *acute*. In Rieser (2010b) it is shown that the two agents resolve the underspecification in different ways, the Follower prefers the cylindric shape version which leads to a local inconsistency in the dialogue. The strategy of using underspecification techniques is different from the one followed in this paper which attributes priority to verbal information and investigates how the gesture "catches up" with it.

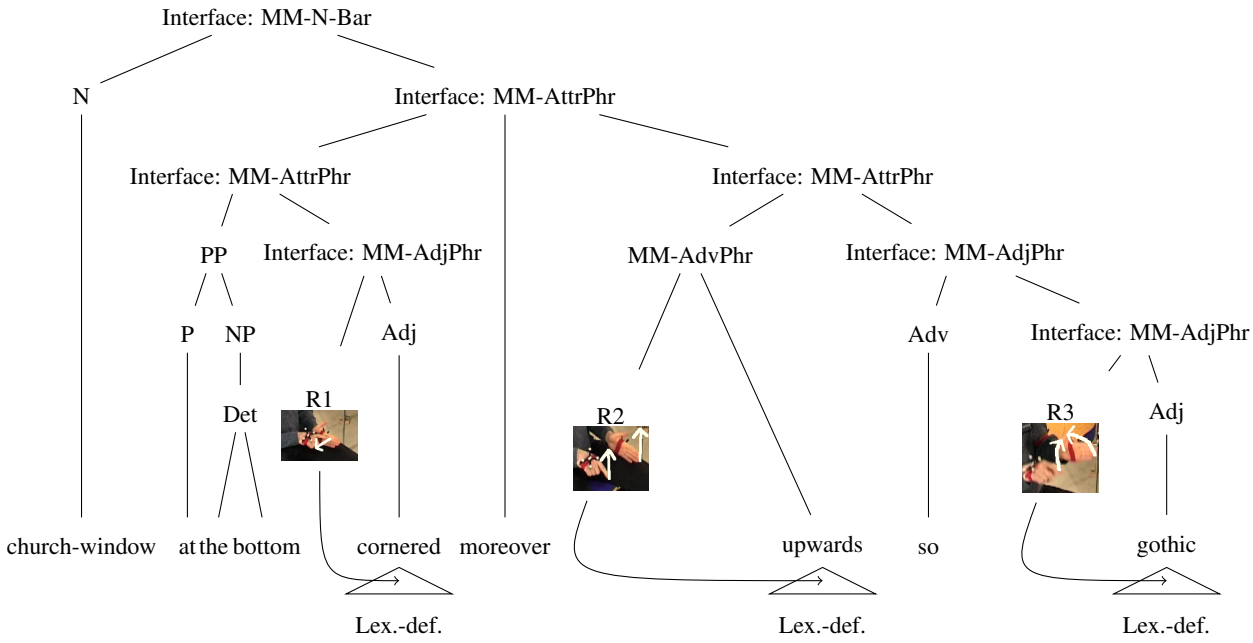

Figure 7: Schedule for the MM contributions and their interfaces. The interface points are at the level of lexical semantics and lead to MM phrases.

feature and cluster layers. The 0-dimensional entities represented by demonstrations have no spatial extension (App., Fig. 11). Furthermore, there are two-dimensional entities like rectangles, squares and so on (App., Fig. 12), followed in complexity by three-dimensional entities such as cuboids (App., Fig. 13). An interesting empirical fact is that we get composites of *n*-dimensional entities. The Router's gestures R1-R3, for example, form a composite of an object's left side, a single horizontal line and a composite line consisting of a vertical part and the bend. The composite depicts partial information about the upper part of a gothic church-window. There are many composites in the V5 datum, the functionally most conspicuous ones being the following: line touching circle orthogonally from the outside (Fig. 14), horizontal and orthogonal line meeting (Fig. 15), two three-dimensional objects held and set into relation to a third one introduced earlier on (Fig. 16).

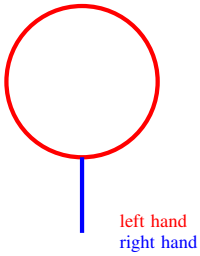

Fig. 14: Composite of two-dimensional and one-dimensional entity. LH holding a round object, RH drawing the path touching the circle.

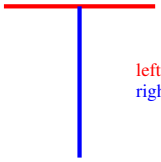

Fig. 15: Composite of two lines: LH modelling line orthogonal to path drawn continuously by RH. The point of contact produced is on the orthogonal line.

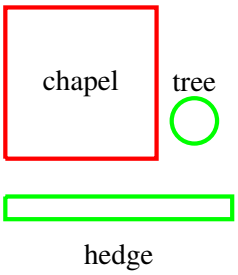

Fig. 16: Composite of 3 three-dimensional entities. LH holding a three dimensional entity (chapel) and RH placing the tree and shaping the hedge in front of the chapel.

5 Statistics of the Typological Grid

Out of the grid data the statistics shown in table 1 was computed. First we have a look at the differences in gesturing between the Router and the Follower. Consider the router's RH: It turns out that the Router mostly uses lines in RH (48%). In the second place come 2D-objects (locations, regions, circles, rectangles etc., 28%), three-dimensional objects in RH are next (prisms, cuboids, cylindroids, spheres etc., 15%) followed by 0-dimensional entities (abstract objects, 9%). The ranking of gestures for the Router's LH is as follows: 0-dimensional entities (7 %) $<$ one-dimensional entities (22%) $<$ three dimensional entities (34%) $<$ two dimensional entities (%

37). In the Follower's RH one-dimensional gestures (lines, 40%) dominate. To a lesser extent he uses two-dimensional (locations, 31%), 0-dimensional (abstract objects, 10%) and three-dimensional (prisms, spheres, 7%) gestures. With his LH he only shapes 0-dimensional objects (abstract objects, 50 %) and three-dimensional ones (prisms, spheres, 50%).

5.1 Interpretation: The Global Picture

Generally speaking, the Router concentrates on depicting routes, regions and locations as well as objects as (parts of) landmarks. Composites consisting of $n \geq 2$ gestures provide the possibility to "hold" the landmarks and sketch the route to them: at the same time both, landmark and route are relationally placed in Router's gesture space (see Figs. 14-16). Interestingly, the Follower sets up his interactive map using one-dimensional gestures most of the time. In other words, he concentrates on representing routes. With both, Router and Follower, the RH is dominating when gesturing (cf. table 1). The Router uses far more two-handed composites than the follower. He populates gesture space with more objects than the Follower does (cf. table 1). As a consequence, his gesture space embodies more information than the Follower's.

5.2 Interpretation: The Role of Features in Detail

The five features, *HandShape*, *BOHDirection*, *PalmDirection*, *WristPosition*, and *WristMovementDirection* are most frequently used by both Router and Follower in their LHs and RHs, respectively. Hence, the annotationally motivated grouping of the features *HandShape*, *BOHDirection* and *PalmDirection* into one *FeatureCluster* at the outset of the typology work (cf. Figs 9 - 13 in App.) gets statistical support. At the same time the large number of *WristPosition* features and *WristLocationMovementDirection* features motivates the set up of clusters for *WristPosition* and *WristMovement* respectively. Both Router and Follower predominantly use their RHs, to be inferred from the greater number of feature clusters there (App., Table 3).

6 Evaluation: Generalizing the Statistical Results for the Grid for the whole SAGA Corpus

The grid material provides a hierarchy of *n*-dimensional gesture categories. How can we use the grid results in establishing the overall SAGA statistics? So far, we can set up the following hypotheses:

1. The grid categories can be used for other data, even for those which do not belong to the SAGA corpus.[5]

2. We know the AVMs for e.g. 0-dimensional objects, one-dimensional objects etc. in detail which we have to look for in annotated material, presupposing, of course, similar annotation conventions. The assumption is that we'll get similar hierarchies for other (even new) data as we have in the grid, i.e. 0-dimensional, one-dimensional etc. objects, depending, of course, on the task. If we had a task dealing with planar objects only, we would presumably have not so many 3-dimensional gestures.

   A cursory investigation of the variables in the rest of the SAGA corpus has shown that this is true. We do indeed have objects of the *n* dimensions established for V5 in most of them. However, most probably, V5 shows the hierarchy in the most explicit way, in other films some layers seem to be missing, e.g. there are no 2- and 3-dimensional entities for the Follower in V1.

3. Concerning the video-datum we started from we can investigate whether the speech-gesture ensembles of the Router and the Follower are structured in a similar way. This is important for research into inter-personal alignment and the role of gesture in interaction.

[5] In reply to a question of one of the reviewers: So far, we are sure that this hypothesis is true of the other video films in the SAGA corpus but we have not tested it with respect to unseen data from different MM corpora. However, the hierarchy extracted from the SAGA data is very general, leading to the assumption that whenever a CP has a kind of linear information he can use a "line"-gesture and similarly for the other dimensions. A restriction concerning generalisation we presumably face could be due to the SAGA domain of concrete n-dimensional objects and routes between them. Whereas we expect that the hierarchy can be used for geometrical objects, CPs discussing sets, functions and λ-terms could well use different gestures.

Table 1: Statistics of the gesture morphology in the typological grid datum ordered by dimensions.

0-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	14	9	5	7	–	–	19	5
Follower	10	22	4	50	–	5	19	27

1-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	75	48	16	22	7	5	103	28
Follower	18	40	–	0	3	0	21	30

2-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	44	28	27	37	–	24	95	25
Follower	14	31	–	0	–	2[1]	16	23

3-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	24	15	25	34	–	67	116	31
Follower	3	7	4	50	–	–	7	10

Mixed-Composites

	OH	TWH	total	total%
Router	–	40	40	11
Follower	–	8	8	10

Totals

	total RH	total LH	total OH Comp.	total TWH Comp.	total
Router	157	73	7	136	373
Follower	45	8	3	15	71

[1] Note: Composites can occur without any corresponding single or one- handed gestures. In that case the composites can't be reduced to single or one-handed gestures. Therefore in this column we have TWH Composites but no LH gesture.

Here we have examples which show that such an investigation makes sense.

4. We can investigate whether the speech-gesture ensembles of other agents and of other pairs of agents in the corpus are structured in a similar way.

5. One can use the partial ontology set up for an *n*-dimensional gesture of the grid for a selected arbitrarily *n*-dimensional one from the rest of the corpus and test whether the former is adequate.

   This has been confirmed for lines and cuboids.

7 Application of Typology: Interfacing Verbal Meaning and Gestural Meaning in MM Dialogue

So far we have seen the following: A dialogue passage with gestures co-occurring with speech and the gesture typology for one complete datum, V5, which gives us a hierarchy of gestures ranging from 0-dimensional entities to *n*-ary composites. The issue we face now is "How can we use the typology in explaining the meaning of multimodal discourse?" We associate the gestures R1-R3 with their descriptions in terms of gesture morphology, both attributes and values. Attributes and values, for example *HandShape*-LH and *Bspread* in R1 are fused into a new attribute where the original value is still "visible" as a suffix. This new attribute is given a stipulated semantic value *side(s1)*

$\wedge$ *of(s1, z)* $\wedge$ *brdth(br, s1)*, in terms of the grid hierarchy, a two-dimensional entity, a side *s1* of something z having a breadth *br*. In a similar manner, the *PathofWrist*-attribute is associated with some length *l1* and the *TwoHandedConfiguration* with a right angle. Of course, not every information contained in the stipulation is derived from the typology but the typology serves as a precondition for the partial ontology.

R1	
HandShape-LH-Bspread	$side(s1) \wedge of(s1, z) \wedge brdth(br, s1)$
PathofWrist-RH-LINE>LINE>LINE	$length(l1, bb) \wedge of(bb, z)$
TwoHandedConfiguration-RFTH>BHA>RFTH>BHA	$right\text{-}angle(s1, ra, bb)$

brdth abbreviates breadth. LH signs a side *s1* of an object *z*. So, there is something z of which *s1* is a side. RH provides a path *l1* emerging from the side *s1* to the right. Both hands produce the right angle of an object *z* shaped by the side *s1* and the planar object *bb*. Intuitively, we have depicted an object *z* with a corner formed by the planar object *bb* and the side *s1*.

R2	
HandShape-LH-Bspread	$side(s1) \wedge of(s1, z) \wedge brdth(br, s1)$
PathofWrist-RH-LINE	$height(h1, ss2)$
WristMovementDirection-RH-MU	$height(h1, ss2)$

LH continues to hold the side *s1*. RH signs the height *h1* of an object *ss2*. The second and the third attribute-value pair provide the same information.

Note that it would be incorrect to identify variables *s1* and *ss2*.

R3	
HandShape-LH-*Bspread*	$side(s1) \wedge of(s1, prr) \wedge brdth(br, s1)$
PathofWrist-RH-*LINE>LINE*	$edge(e1, prr) \wedge edge(e2, prr)$
WristMovementDirection-RH-*MR/MU>MD/MR*	$angle(e1, a, e2) \wedge acute(a)$

The side *s1* is still held by LH. The RH produces two edges *e1* and *e2* of an object *prr* which form an acute angle at the top.

Linking up R1-R3 with the *church-window* Property:

The overall speech-gesture meaning integration can be derived from Fig. 7. We cannot show that in detail here. The final structure of the MM meaning of the N-Bar construction plus the accompanying gestures is:

$window(w) \wedge$ *part-of*$(w, wa) \wedge wall(wa, ch) \wedge church(ch) \wedge$ *lower-part-of*$(lp, w) \wedge$ *middle-part-of*$(mp, w) \wedge$ *upper-part-of*$(up, w) \wedge lp \oplus mp = cub \wedge cuboid(cub) \wedge base(b1, cub) \wedge breadth(br, b1) \wedge side(s1, lp) \wedge$ (*upper-part-of*$(up, w) \Leftrightarrow (prism(up) \wedge pointed(up) \wedge acute(up))) \wedge$ $\underline{at(w, lp)} \wedge \underline{\text{lower-part-of}(lp, w)} \wedge \underline{side(s12, lp)} \wedge$ $\underline{\text{right-angle}(b1, ra, s12)} \wedge \underline{length(l1, s12)} \wedge \underline{height(h1, ss2)}$.

Comparing this result with the lexicon-entry for *church-window* introduced in 3.3 shows that the lexicon-entry remained consistent. Only the portion marked by underlining is additional information. It contains the information needed for the lower part of the church-window. So we see that iconic gestures can highlight which aspect of word meaning is intended from the CP's point of view, thus supporting parts of the more analytic lexical definition. This is especially clear from the disjunction: Only the "pointed" and "acute" version is consistent with the gesturing.

Acknowledgements

Work on this paper has been carried out in the project B1, *Speech-gesture Alignment*, of the CRC *Alignment in Communication*, Bielefeld University. Support by the German Research Foundation is gratefully acknowledged. We also want to express our thanks to our fellow researchers Kirsten Bergmann, Stefan Kopp and Andy Lücking, as well as to three anonymous SEMdial reviewers.

References

Bergmann, K., Fröhlich, C., Hahn, F., Kopp, St., Lücking, A. and Rieser, H. 2007. *Wegbeschreibungsexperiment: Grobannotationsschema.* Bielefeld Univ.

Bergmann, K., Damm, O., Fröhlich, Hahn, F., Kopp, St., Lücking, A., Rieser, H. and Thomas, N. 2008. *Annotationsmanual zur Gestenmorphologie* Bielefeld Univ.

Ekman, P. and Friesen, W. 1969. The repertoire of nonverbal behavior: categories, origins, usage and coding. *Semiotica 1*, pp. 49-98

Kendon, A. 2004. *Gesture: Visible Action as Utterance.* CUP.

Kita, S. (Ed.) 2003. *Pointing: Where Language, Culture and Cognition Meet.* Lawr. Erlb.

Lücking, A., Bergmann, K., Hahn, F., Kopp, St., Rieser, H. 2010. The Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). Submitted for LREC 2010.

McNeill, D. 1992. *Hand and Mind. What Gestures Reveal About Thought.* Univ. of Chic. Press

Ginzburg, J.: 2010. *The Interactive Stance. Meaning for Conversation*. MS, King's College London, (to appear).

Poesio, M. and Rieser, H.: 2010. An Incremental Model of Anaphora and Reference Resolution Based on Resource Situations. (Submitted to DD, SI on Incrementality).

Poggi, I. 2002. From a Typology of Gestures to a Procedure for Gesture Production. In: Wachsmuth, I. and Sowa, T. (eds.), *Gesture and Sign Language in Human-Computer Interaction*. LNAI 2298, pp. 158-168

Rieser, H. 2004. Pointing in Dialogue. In *Proceedings of Catalog* '04, pp. 93-101.

Rieser, H.: 2010a. On Factoring out a Gesture Typology from the Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). In Kopp, St. and Wachsmuth, I. (Eds.), *Proceedings of GW 2009,* Springer: Berlin, Heidelberg, pp. 47-60.

Rieser, H.: 2010b. How to Disagree on a Church Window's Shape Using Gesture. In: *Proceedings of SEMdial 2010.*

Rieser, H. and Poesio, M.: 2009. Anaphora and Direct Reference. Empirical Evidence from Pointing. In: P*roceedings of DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue.* Stockholm, Sweden, pp. 35-43.

Appendix

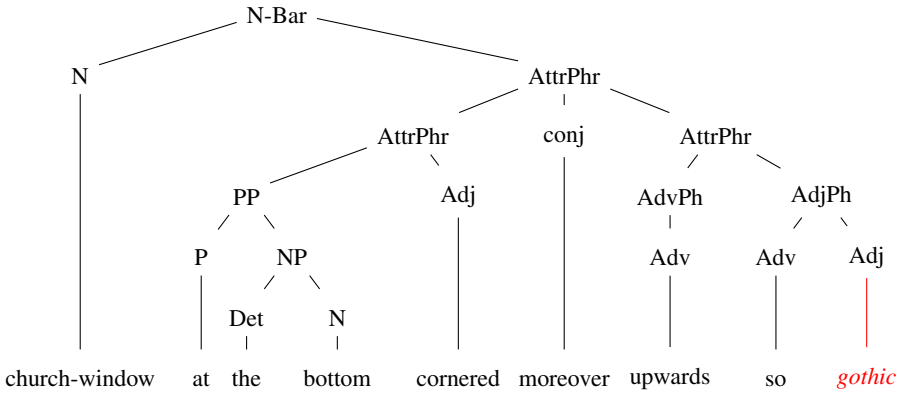

Figure 4: Represents the completed NP with the Follower's *gothic*, triggered by the Router's gesture, inserted. Observe that *in toto* we have a cooperatively produced AdjPh.

Table 2: Non-zero-valued features used in the typological grid.

Features	**Router**		**Follower**	
	LH	RH	LH	RH
PathOfWrist	3	3	2	3
WLMDirection	10	14	2	9
WLMRepetition	2	3	1	2
WristDistance	2	3	2	2
WristPosition	9	13	11	18
BackOfHandDirection	11	11	4	10
BOHMDirection	5	5	1	1
BOHMDRepition	1	1	1	1
PathOfBOH	3	3	1	1
PalmDirection	14	13	5	12
PDMDirection	3	3	1	1
PDMRepitition	1	1	1	1
PathOfPalm	2	2	1	1
HandShape	22	22	4	12
HSMRepitition	1	2	1	1
HSMDirection	1	6	1	1
PathOfHandshape	1	2	1	1
TemporalSequence	1	2	1	1
TWH Features	both hands		both hands	
TWH-Configuration	13		9	
TWH-Movement	6		3	

Table 3: Clusters used in the typological grid.

Cluster	**Router**		**Follower**	
	LH	RH	LH	RH
PMovement	3	2	1	1
HSMovement	1	4	1	1
BOHMovement	4	6	1	1
FeatureCluster	71	109	6	33
WristMovement	26	50	2	16
WristPosition	34	52	12	30
TWH Cluster	both hands		both hands	
TWH-Cluster	35		19	

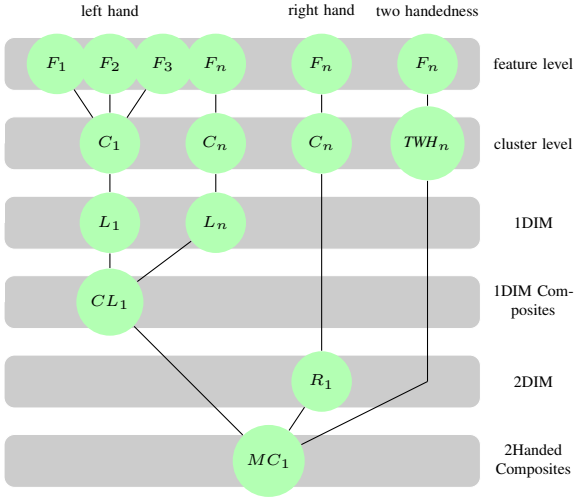

Figure 17: Section of gesture hierarchy (simplified).

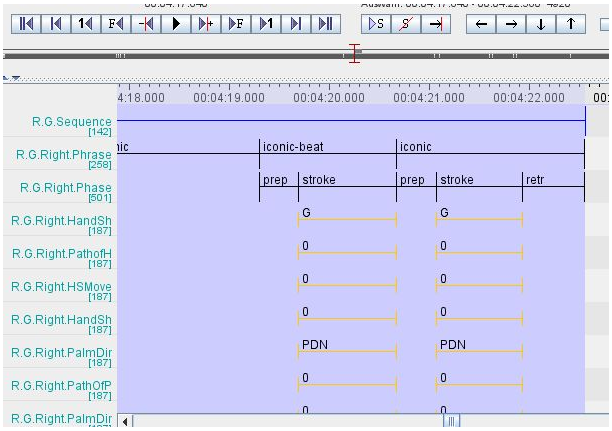

Figure 3: Example of the gesture annotation. It partially represents R1 and R2 from Fig. 2.

R-Line-RH		
R-FeatureCluster-RH-1a	*R-FeatureCluster-RH-1a-cat*	
	HandShape	*G*
	PalmDirection	PDN
	BackOfHandDirection	BAB
R-FeatureCluster-RH-2a	*WristMovement-RH-1a-cat*	
	PathofWrist	*Line>Line>Line*
	WristLocationMovementDirection	*MR>ML >MR*
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3a	*WristPosition-RH-1a-cat*	
	WristPos	CLW
	WristPosDist	DEK

HandShape	*G*

Figure 9: *Feature HandShape* and its value *G*.

Figure 10: Gesture representing a horizontal line.

R-Direction-G212-RH		
R-FeatureCluster-RH-1a	*R-FeatureCluster-RH-1a-cat*	
	HandShape	*G*
	PalmDirection	PDN
	BackOfHandDirection	BAB/BTL
R-FeatureCluster-RH-2a	*WristMovement-RH-1-cat*	
	PathofWrist	$\emptyset$
	WristLocationMovementDirection	$\emptyset$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3a	*WristPosition-RH-1-cat*	
	WristPos	CUP
	WristPosDist	DEK

Figure 11: 0-dimensional entity direction.

R-Rectangle-LH		
R-FeatureCluster-LH-1c	*R-FeatureCluster-LH-1a-cat*	
	HandShape	*B spread*
	PalmDirection	*PTR*
	BackOfHandDirection	BAB
R-FeatureCluster-LH-2a	*WristMovement-LH-1a-cat*	
	PathofWrist	$\emptyset$
	WristLocationMovementDirection	$\emptyset$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-LH-3c	*WristPosition-LH-1a-cat*	
	WristPos	CLL
	WristPosDist	DEK

Figure 12: 2-dimensional entity rectangle or side.

R-Cuboid-G103-RH		
R-FeatureCluster-RH-1c	*R-FeatureCluster-RH-1az-cat*	
	HandShape	*small C*
	PalmDirection	*PAB*
	BackOfHandDirection	*BUP*
R-FeatureCluster-RH-2a	*WristMovement-RH-1ax-cat*	
	PathofWrist	$LINE > LINE > LINE$
	WristLocationMovementDirection	$MF > MB > MF$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3c	*WristPosition-RH-1q-cat*	
	WristPos	*CC*
	WristPosDist	$DCE > DEK$

Figure 13: 3-dimensional entity cuboid.