Aligned Iconic Gesture in Different Strata of MM* Route-description Dialogue

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

Abstract

This paper deals mainly with iconic gesture in two-agent route description dialogue and focuses largely on the interface of word semantics and gesture. The modelling tools used come from formal semantics and pragmatics. The empirical background of the study is a partly annotated corpus of ca 5.000 gestures collected in the Bielefeld Speech-and-Gesture-Alignment Corpus (SAGA). The approach taken is entirely new: an interface comprising word meaning and gesture meaning is constructed, the point of contact being the temporal overlap between gesture and speech in the annotated data. Gesture meaning is computed *via* a mapping *rep* from the set of annotation predicates onto a meaning representation. There is a discussion concerning the trade-off between context-free *vs.* context-dependent word meaning and gesture meaning. The interfaced speech-gesture meaning is represented in a dynamic semantics format easily grafted on a formal syntax fragment.

*MM stands for *multi-modal*.

1 Introduction[1]

It is well known that gestures of agents are ubiquitous in dialogue (cf. McNeill (ed. (2000)), Kita (ed. (2003)) but not where it can be placed in dialogue and what then will be its function there. Judged by experience with corpus data and the gesture folklore there is little doubt that there is pointing to objects in context (cf. Rieser (2008)) and that properties such as rectangularity can in a way be indicated by gesture. However, is there something more definite that can be said? As far as we know there has been no work on MM dialogue so far investigating these matters on a more principled basis. Below it will be shown that gestures can go into different structural positions in dialogue, exhibiting different meanings and functions. Even if we rely on a fairly large corpus of multi-modal dialogue, the (Bielefeld University) SAGA corpus elicited in a strictly controlled VR experiment, comprising roughly 5.000 gestures, the evidence presented here cannot be conclusive. There might still be other functions and most plausibly, there are. Nevertheless, we claim that the findings we show and explain are prototypical for natural MM dialogue. So, in section 1 we will provide an overview on structural positions observed for gestures in MM dialogue. Ch. 2 will deal with a binding problem of some sort, namely, how gesture

[1] In this paper only literature is quoted which has been evaluated as relevant for its methodological concerns, which is largely formal theory building. So, some readers might miss their favourite papers. Thanks go to three anonymous reviewers who raised a lot of interesting issues. Some of their arguments are taken up below, space permitting. Sometimes I will refer to a reviewer's (abbr. as rev. *n*'s) remark.

Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue, June 2–4, 2008, London, U.K.

information can be 'bound'[2] to speech information. Ch. 3 will deal with the interface of gesture meaning and verbal meaning, restricted to word meaning, and there will be a brief discussion of the methodological problems with this approach in ch. 4.

2 Overview on Structural Positions Observed for Gesture in Dialogue

As an introduction to the function of gesture in dialogue, we set out with a naïve methodology and provide prototypical speech-gesture occurrences. We might view these as instances of ratings leading to systematic annotation, i.e. we first do speech-gesture pairings in a naïve way; as a consequence, the total meaning of speech plus gesture is given in the short descriptions. Of course, the coordination of speech and gesture information is a major *explicandum* of this paper, so this introductory perspective will be given up in sections 2-4, where the ontological status of speech meaning and gesture meaning is discussed and the speech-gesture interface is the central issue. The stills in fig. 1 below show the stroke positions of iconic gestures; it should be kept in mind that gestures are incomplete and even non-standard in various ways and provide partial information at best. So, in interpreting gestures we have to assume top-down Gestaltist processes at work. (a) is an oval gesture accompanying the description of a sculpture indicating part of the concrete basis for the sculpture, (b) presents a gesture indicating the two towers of a church, in (c) the route follower imitates the router's gesture indicating the U-shape of the town hall, (d) has an other-correction carried out by a router's gesture, (e) has a two-handed gesture which depicts a situation containing a chapel and a tree. (a) and (b) are routers gestures, (d) has interaction resting on gestures functioning like turns. Fig. 2 gives a summary of these findings, indicating the various functions of gestures.The data in Fig. 1 are related to Fig. 2 as follows: The gesture in still (a) is related to word semantics, the one in (b) to the semantics of an NP-constituent, in (c) a gesture goes proxy for a propositional content which gets acknowledged, (d) shows that a gesture is used in a next turn repair, in (e), finally, the right hand models a tree while the left hand indicates the location of the tree beside a chapel.

The example 1 discussed below (cf. fig. 4) will deal in some detail with the extension of word meaning by gestural meaning.

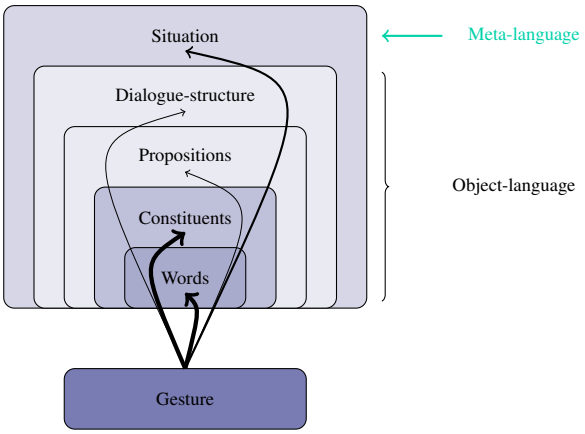


Figure 2: Summary of observations concerning the structural positions and functions of gestures in MM dialogue.[3]

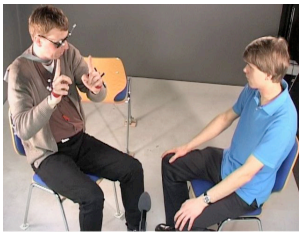
3 A Binding Problem Involving two Representations: How Speech and Gesture Information Are Interfaced

The description of gesture functions provided above may seem fairly convincing, however, we are interested in answering the following questions (a) Do iconic gestures have meaning? (b) Given that they do, how does their meaning interact with verbal meaning? Question (a) has been answered positively in the tradition of semiotic research going back at least to Ch. S. Peirce and carried on in the gesture context by McNeill, Cassell and others. Even if it is difficult to tell how exactly one can provide meanings for gestures on the basis of gesture tokens, we assume here that the representation of gesture mean-

[2]The notion of binding used here is taken from neurobiology and vision research. There is little doubt that the logical notion of operator binding can also be related to these more fundamental notions.

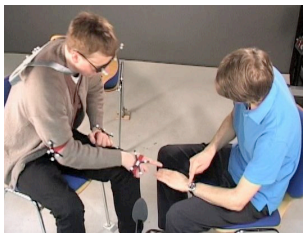
[3]Rev. 1 did not approve of the meta-language label used here. The point is simply that there is no *a priori* argument for putting the formally reconstructed gesture meaning into either the object language or into the specification of the model used. Intuitively, the information of some bi-manual non-symmetric gestures is better placed into a model's definition of domains. One could even start with the hypothesis that gestures generally depict partial models and do not go into the object language at all but investigation of this research line has to wait for another paper.

(a) The concrete sculpture base



(b) Two towers of a church

(c) The U-shape of the town hall



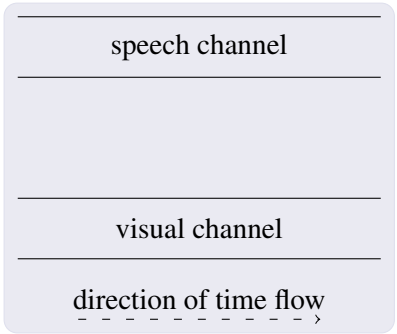
(d) Other correction of follower's gesture

(e) The chapel and a neighbouring tree

Figure 1: Stills showing structural positions of iconic gestures in MM dialogue.

ing can be given in much the same way as for verbal tokens. As a first orientation, assume that gesture meaning behaves functionally like the meaning of deictic expressions. Turning to (b), we will gradually develop a workable schema for a speech-gesture interface below. Starting from the folklore assumption that speech and gesture sit on different channels, we get the picture in Fig. 3, with two channels running in parallel and no interaction specified between speech and gesture. This is meant to serve as our didactic starting point to be modified in stepwise fashion.

Figure 3: Speech channel and visual channel running in parallel



However, there must be an interaction of some sort, since non-lexicalized iconic gestures cannot provide a semantics on their own, so the argument goes in some of the literature (see Kopp et al. (2004) and Lascarides and Stone (2006)). Now the interaction could be of different sorts, e.g. it might be the case that (a1) we can construct some total object language meaning out of the two sorts of meanings or that (a2) we consider one type of meaning as a context to interpret the other type of meaning. An extreme version of (a2) delegates gesture meaning to the context, in particular, to the specification of the model, over which the object language expression is interpreted. So, the function of the MM meaning produced or observed is split, some part goes into the object language and the other into the meta-language. Fig. 2 above indicates that data tell us, when to regard gesture meaning as part of the object language and when to consider it as part of the model. (a2) has as a consequence that one considers only models which satisfy the information provided by the gesture. As a matter of fact, we get most of the information needed for our design decision for an object language (a1) or a meta-language (a2) solution from the annotation depicted in Fig. 4.

[4]The annotation follows two working manuals (Bergmann *et al.* (2007b) for practices and Bergmann *et al.* (2008) for handshapes). The six researchers annotating have been trained over some month on ample raw data; their rate of agreement was

Start Time	End Time
0:39.170	0:41.780
Right.Handshape.Shape	large C
Right.Path.of.Handshape	0
Right.Handshape.Movement.Direction	0
Right.Handshape.Movement.Repetition	0
Right.Palm.Direction	PAB
Right.Path.of.Palm.Direction	0
Right.Palm.Direction.Movement.Direction	0
Right.Palm.Direction.Movement.Repetition	0
Right.Back.of.Hand.Direction	BAB/BUP
Right.Path.of.Back.of.Hand.Direction	0
Right.Back.of.Hand.Direction.Movement.Direction	0
Right.Back.of.Hand.Direction.Movement.Repetition	0
Right.Path.of.Wrist.Location	ARC
Right.Wrist.Location.Movement.Direction	$MR > MF$
Right.Wrist.Location.Movement.Repetition	0
Right.Extent	medium
Right.Temporal.Sequence	0
Left.Handshape.Shape	large C
Left.Path.of.Handshape	0
Left.Handshape.Movement.Direction	0
Left.Handshape.Movement.Repetition	0
Left.Palm.Direction	PAB
Left.Path.of.Palm.Direction	0
Left.Palm.Direction.Movement.Direction	0
Left.Palm.Direction.Movement.Repetition	0
Left.Back.of.Hand.Direction	BAB/BUP
Left.Path.of.Back.of.Hand.Direction	0
Left.Back.of.Hand.Direction.Movement.Direction	0
Left.Back.of.Hand.Direction.Movement.Repetition	0
Left.Path.of.Wrist.Location	ARC
Left.Wrist.Location.Movement.Direction	$ML > MF$
Left.Wrist.Location.Movement.Repetition	0
Left.Extent	medium
Left.Temporal.Sequence	0
Two.Handed.Configuration	$FTT > BHA$
Movement.realtive.to.other.hand	mirror-sagittal

Figure 4: Annotation of example: router's contribution (1) *die Skulptur die die hat 'n Betonsockel / the sculpture it it has a concrete base* [4]

The annotation specifies features and functions of the router's left and right hand, both, on a more global level (the so-called practices like indexing, shaping or grasping giving the global function of the gesture) and on a more fine-grained level which captures the postures of both hands, their parts (palm, back-of-hand, wrist etc.) and their respective movements (left, right, forward etc.). However, the most important thing in the annotation grid is that it maps speech and gesture onto a time line; hence, we can see which speech occurrences overlap with which gesture occurrences. Intuitively, we consider the flowing time as more basic information by help of which speech and gesture events can communicate. Communication among events on different channels is brought about or even caused by temporal synchronization of inputs. This is the concept of binding referred to above. There are several supporting arguments for the binding of gesture meaning to verbal meaning and vice versa:

(1) McNeill (1995, pp. 26-31) considers the stroke information as the carrier of the central semantic and pragmatic information of the gesture. It is in turn tied to the corresponding constituent's stressed syllable or, as we prefer to put it, 'aligned', i.e. synchronized with it. See (Lücking, Rieser, Stegmann (2004)) for experimental evidence.

Supporting arguments (2) and (3) operate on a neuro-information level, (2) concerns vision and (3) cognition in general:

(2) Neuro-biological research on vision is devoted to the so-called binding problem, the dominant model entertained being the time-coding model: the temporal synchronization of the stimuli is the decisive mechanism for integration (Detel (2007), p. 33, translated by the author).

(3) *[Likewise] events that coincide in time are interpreted with greater probability as [being] related than events separated in time* (Singer (1999)).

(4) Finally, from a Gestaltist perspective, rules of grouping and proximity apply.

We cannot enter the difficult problem of neural representation here but will stick to the tools of linguistics and philosophy of language. A rough picture illustrating the information flow of synchronized (aligned) information still using the channel concept is provided in fig. 5. It shows that if there is temporal alignment among events from the different channels, then information from the gesture channel is coordinated with the information from the verbal channel by binding.

4 Interface of Gesture Meaning and Verbal Meaning

We now follow the research strategy a1 introduced in sect. 3. From fig. 5 we see that the following is needed to model the interaction of gesture and

tested in several studies and amounted to 80% in most cases. The temporal boundaries used in example 4 were rated.

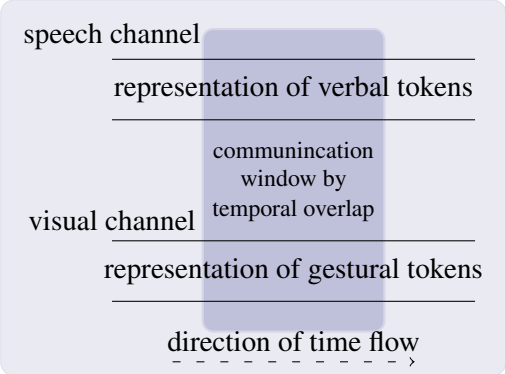


Figure 5: Binding in between the gestural and the verbal channel depending on time synchrony.

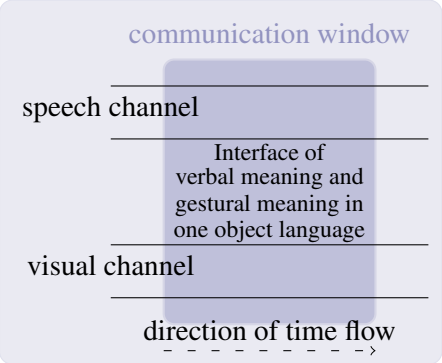


Figure 6: Interface in the communication window established in between the channels.

speech: a representation of (a) the verbal information, of (b) the gesture information compatible with 'Marr structures' (Marr (1982)), and (c) a point of contact for linking the different types of information. (a), (b), and (c) can be achieved using type logics or unification. Gesture information is drawn from the descriptive predicates and values of the fine-grained annotation. For reasons of simplicity we can regard the verbal information as the function operating on the information of the gesture level. However, both must be conceived of as dynamic, due to the direction of time flow on both channels. These inherent constraints can be met by several types of Dynamic Semantics, *inter alia* classical DRT, SDRT, Muskens LDG and PTT, all these add information updating already existing information. The point of contact between the verbal and the gestural level is provided by the window given by temporal overlap (see fig.5), hence temporal synchrony is what matters (i.e. regarded as a necessary condition).[5] The methodological grid now emerging is shown in fig. 6: verbal information and gesture information are interfaced and establish together the context for new information to be integrated. Integration will be anticipated by open slots in the already existing information.

We now specify the procedures for the annotation example in some more detail and concentrate on extracting the semantics of the gesture out of the annotation predicates; the representation of the dynamic semantics of the verbal contribution *die Skulptur die die hat 'n Betonsockel / the sculpture it it has a concrete base* is far from trivial, but we gloss over it here. In the MM example we have the temporal overlap between *Betonsockel/concrete* base and the gesture shown in fig. 1 (a). So, the necessary condition for a fusion of the verbal meaning and the gestural meaning is given, meeting hypotheses (2) - (4) in sec. 3. What do the hands involved sign or inscribe? Here we consider only the relevant parameters in the stroke phase, meeting in particular McNeill's hypothesis (1). The parameters and their values are represented as typed feature structures with types written in italics and standard attribute value pairs <attribute value> used (fig.7).

The matrices show postures of the router's left and right hand as well as two-handed postures. In methodological terms, the annotation predicates constitute the observational language which provides the foundation for our theoretical terms, i.e. the semantic predicates. Figure 8 shows the volume or space shaped by both hands using the annotation predicates as labels.

So, what do both hands depict? Looking at the *R.G.Left* and the *R.G.Right* information, we see that the wrists follow ARC paths. In the beginning, fingers and thumbs touch ($= FTT$), but they separate immediately ($= \neg FTT$). The C-shapes on both hands provide us with a dense series of verticality informations. They also indicate some of the information of a top and a bottom (marked by the top- and bottom-curves of C respectively). $ML > MF$

[5] Rev. 1 does not agree with this assumption, whereas rev. 3 finds it trivial. However, temporal relation of events is the most conspicuous information we can get hold of in the observational data. The ultimate evidence is, of course, a consistent formal model, cf. the remarks in section 2 *A Binding Problem etc*.

$$\begin{bmatrix} \textit{Both hands} & \\ \begin{bmatrix} \textit{R.G.Left} & \\ \text{HandShape} & \textit{loose C} \\ \text{Palm Direction} & \textit{PAB} \\ \text{BackofHand} & \textit{BAB/BUP} \\ \text{PathofWrist} & \textit{ARC} \\ \text{WristLocation} & ML > MF \end{bmatrix} & \begin{bmatrix} \textit{R.G.Right} & \\ \text{HandShape} & \textit{loose C} \\ \text{Palm Direction} & \textit{PAB} \\ \text{BackofHand} & \textit{BAB/BUP} \\ \text{PathofWrist} & \textit{ARC} \\ \text{WristLocation} & MR > MF \end{bmatrix} \\ \text{Two-handedConfiguration} & FTT > \neg FTT \\ \text{Movement relative to other hand} & \textit{Mirror-sagittal} \end{bmatrix}$$

Figure 7: Typed feature structures for some of the information provided in the annotation of fig. 4.

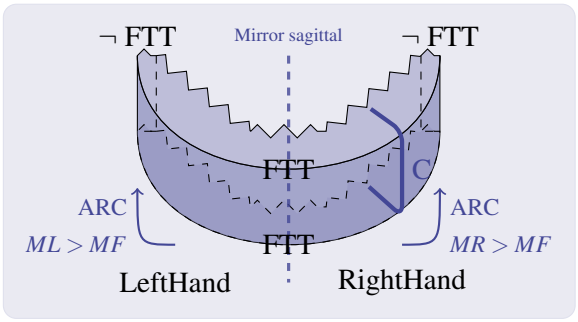


Figure 8: Typed feature structures for some of the information provided in the annotation of fig. 4.

(left forward) and $MR > MF$ (right forward) trace the extent of the curved lines of the sectors bounded by ARC lines. PalmDirection values and BackofHand values follow from the ARC and the WristLocation predicates. We have wrist movements to the left and the right. Finally, Mirror-sagittal shows symmetric extent of the left and the right segment from the router's perspective. What we do now is provide a mapping from the descriptive annotation predicates into a semantic domain. It must specify the depictional value of the gestures and also fix their iconic functions. Thus, the notion of 'similarity' is eliminated *via* a semantic interpretation. Mappings like these have been argued for in (Rieser (2004)) and in (Lascarides and Stone (2006)). We assume a conventional basis for these mappings in Grice's or Lewis' sense, which might depend on a class of contexts: obviously, there must be a reason why we understand gestures and can reliably annotate occurrences of them. The function *rep* indicates representation. *rep* goes from the set of annotation predicates into open formulas. So, the denotation for gestures is provided via translation.[6]

(2) (a) $rep(\text{HandShape } looseC) = hight(x,u) \wedge top(t,u) \wedge bottom(b,u)$

(b) $rep(\text{PathofWrist } ARC) = curved\text{-}side(s,u)$

(c) $rep(\text{WristLocat } ML > MF) = curved\text{-}side\text{-}left(sl,u,router)$

(d) $rep(\text{WristLocat } MR > MF) = curved\text{-}side\text{-}right(sr,u,router)$

(e) $rep(\text{Movement relative to other hand } Mirror\text{-}sagittal) = part(p1,u) \wedge part(p2,u) \wedge (p1 \neq p2) \wedge (p1 \otimes p2) = u$ [7]

In (c) and (d) the routers perspective is coded because of the direction information requiring a *Bühler origo*. The function *rep* induces a mapping from the gesture space GS onto a semantic space SGS.

5 Canonical Word Meaning and How it Can be Extended Using Gesture Information

For purposes of illustration we now assume the following word meaning for *concrete base/Betonsockel*:

(3) $concretebase(x) := support(x,y) \wedge made\text{-}of\text{-}concrete(x) \wedge rigid(x) \wedge object(y) \wedge (x \neq y) \wedge hight(h,x) \wedge side(s,x) \wedge top(t,x) \wedge bottom(b,x).$

[6] (Taking up remarks by all three reviewers). Two problems should be mentioned here. The mapping *rep* is based on observations. It should doubtlessly be backed by statistical data, which are, as yet, not available. Another interesting point is which formal language should be used to represent the gesture meaning. Here I'm still experimenting (cf. also foot-note 7 on fusion). Looking into versions of Mereotopology (see Casati and Varsi (1999) for an overview), I find, that the standard systems available are not strong enough to represent indexical spatial gestures.

[7] The conjunct $(p1 \otimes p2) = u$ is read as 'parts $p1 \otimes p2$ fused yield the whole u', a suggestion I owe to A. Lücking.

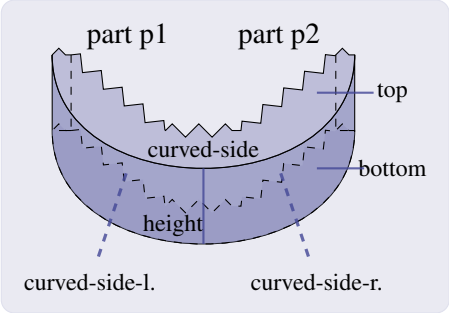


Figure 9: Semantic space SGS induced by gesture space GS, curved lines indicating partiality

So, a concrete base is a support x for an object y iff[8] it is made of concrete, has hight, a side, a top and a bottom. Now, (3) may well be too rich a word meaning for *concrete base/Betonsockel*. So we reduce it and provide an open slot *gest* for the conjunction of the contextual gestural information coded by λ-abstraction in the following way:

(4) $\lambda gest(concretebase(x) := support(x,y) \wedge made\text{-}of\text{-}concrete(x) \wedge object(y) \wedge rigid(x) \wedge (x \neq y) \wedge gest)$

The idea is to model binding between the verbal meaning and the gesture meaning using functional application of (4) for the right-hand-side of (2) as the argument. Hence (4) acts as a context for the gesture information and consumes it. We get

(5) $concretebase(x) := support(x,y) \wedge made\text{-}of\text{-}concrete(x) \wedge object(y) \wedge rigid(x) \wedge (x \neq y) \wedge hight(z,u) \wedge top(t,u) \wedge bottom(b,u) \wedge curved\text{-}side(s,u) \wedge curved\text{-}side\text{-}left(sl,u,router) \wedge curved\text{-}side\text{-}right(sr,u,router) \wedge part(p1,u) \wedge part(p2,u) \wedge (p1 \neq p2) \wedge (p1 \otimes p2) = u.$

What we want to show is:

(a) Contextually, we can do with a minimal word meaning for *concrete base/Betonsockel* consisting of concrete support x for an object y.

(b) Word meaning and gesture meaning interact in context due to temporal binding.

(c) The interface between word meaning and gesture meaning gives us the contextually needed MM meaning which will be more specific than the typical context-free word meaning, and, above all, depend on the situated perspective of the router.

Before we can succeed with (a) - (c), however, we have to deal with the alignment of the objects involved in gesture and speech. Observe that the variables for the logical subjects in (5), x and u, will, as a rule, denote different objects and only contingently refer to the same thing. Intuitively, however, words and gestures in the interface window are about the same object. So, we can formulate the following alignment-of-variables convention:

(6) If words and gestures are about the same object, the same variable must be used for it in the specification of the MM content.

Observing (6) we get an intuitively adequate word meaning.[9]

6 Discussion

In this paper we have only treated the "gesture meaning specifies word meaning" case. We want to take up a few problems. They concern in turn: (1) The reliability of the mapping (2); (2) Dynamic Semantics for lexical information and the embedding of word meaning into the meaning of example (1); (3) Options for reconstructing the relation of word meaning and gestural meaning. Ad (1): Reliability considerations are of course important here, since interpretation and interface construction depend on them. From observation we know that C typically has the function indicated and the same holds for the wrist movements. A slightly different argument in support of the mapping is that one would not find a natural model for example (1) which does not exhibit the gesture semantics indicated. Ad (2): Observe that we can use a dynamic semantics format for the lexical entries. In (5), e.g., we can establish equivalence between two DRSs.We cannot go into matters of establishing a full syntax-semantics interface here, so a few hints must suffice. (7) shows a representation of example (1) in a Muskens LDG format (Muskens

[8]The *iff*-condition will, as a rule, be too strong for word meanings. It is here chosen for reasons of simplicity and perspicuity.

[9](Taking up rev. 1's remarks): All iconic gestures will get a representation using the mapping *rep*. The step from (4) to (5) is computed as described above, modelling *binding between the verbal meaning and the gesture meaning using functional application of (4) for the right-hand-side of (2) as argument*.

(1996)), based on an LTAG representation for reasons of getting at incrementality:

(7) $[x|concretebase(x); \iota y[\ |sculpture(y)] = it; have(it,x)].$

Sticking to the format of explicit definition, we can substitute the right side of expression (5) suitably represented for *concrete base*(x). Hence, intuitively, we will get suitable derivation- and entailment-relations. Ad (3): You may have noticed that the word meaning in (3) does not fully specify the shape of the figure's tops and bottoms. Assume, we add *elliptical(t)* and *elliptical(b)* in order to provide the missing information. Then we run into a problem with (5), since (5) only partially provides the information of an extended (3). It turns out that we encounter a Gestalt regularity here, the principle of Prägnanz (*minimum principle*) being at stake. Perhaps we could solve cases like this one using abduction but it is not trivial to do this. Another Gestalt issue is that gestural movements are not precise in the geometry sense. We leave these topics for a methodology paper.

Acknowledgements

The work reported in this paper has been supported by the German Research Foundation (Project B1, *Speech-Gesture Alignment*, CRC *Alignment in Communication*, Bielefeld University) which is gratefully acknowledged. Thanks go to my co-workers Kirsten Bergmann, Andy Lücking, Florian Hahn and Stefan Kopp for discussion and support.

References

Bergmann, K. and Rieser, H. 2007. Discussion of A. Lascarides and M. Stone's Example (1) from their *Formal Semantics for Iconic Gesture.* Workshop-contribution, Bielefeld Univ., June 2007

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

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

Casati, R. and Varsi, A. C. 1999. *Parts and Places. The Structures of Spatial Representation.* The MIT Press: Cambr., Mass.

Detel, W. 2007. *Grundkurs Philosophie, Bd. 4, Erkenntnis- und Wissenschaftstheorie.* Reclam: Stuttgart.

Kita, S. (ed.) 2003. *Pointing. Where Language, Culture, and Cognition Meet.* Erlbaum: London.

Kopp, St., Bergmann, K. and Wachsmuth, I. 2008. *Multimodal Communication From Multimodal Thinking – Towards an Integrated Model of Speech and Gesture Production* In *International Journal of Semantic Computing* (in print).

Kopp, St., Tepper, P. and Cassell, J. 2004. *Towards integrated micro-planning of language and iconic gesture for multimodal output.* In *Proceedings of ICMI.*

Lascarides, A. and Stone, M. 2006. *Formal Semantics for Iconic Gesture.* In *Proceedings of Brandial*. Potsdam University.

Lücking A., Rieser, H. and Stegmann, J. 2004. *Statistical support for the study of structures in multimodal dialogue.* In *Proceedings of Catalog* 04, pp. 56-64

Lücking, A., Pfeiffer, Th. and Rieser, H. 2008. *Pointing Reconsidered.* Submitted

Marr, D. 1982. *Vision.* San Francisco: Freeman

McNeill, D. 1995. *Hand and Mind. What Gestures Reveal about Thought.* UCP: Chicago and London.

McNeill, D. (ed.). 2000. Language and gesture. CUP.

Muskens, G. 1996. *Combining Montague Semantics and Discourse Representation.* In: *Linguistics and Philosophy*, 19, pp. 143-186.

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

Rieser, H. 2005. *Pointing and Grasping in Concert. With an Encore on Salience.* In: Stede et al. (eds.), *Salience in Discourse: Multidisciplinary Approaches to Discourse* 2005. pp. 129-139

Rieser, H. 2007. *Multimodal action: Demonstration and reference.* In: *IPA Abstracts.* Göteborg, Sweden, pp. 148-149

Rieser, H., Kopp, St. and Wachsmuth, I. 2007. *Speech-Gesture Alignment.* In: *ISGS Abstracts, Integrating Gestures.* Northwestern University, Evanston, Chicago, pp. 25-27

Rieser, H. and Staudacher, M. 2008. *SDRT and Multimodal Situated Communication.* Submitted

Singer, W. 1999. *Neural Synchrony: A Versatile Code for the Definition of Relations.* In *Neuron*, Vol. 24, pp. 49-65.