Gesture meaning needs speech meaning to denote – A case of speech-gesture meaning interaction

Insa Lawler
Department of Philosophy
University of Duisburg-Essen
insa.lawler@uni-due.de

Florian Hahn and Hannes Rieser
Faculty for Linguistics and Literary Studies
Bielefeld University
{fhahn2,hannes.rieser}
@uni-bielefeld.de

Abstract

We deal with a yet untreated issue in debates about linguistic interaction, namely a particular multi-modal dimension of meaning-dependence. We argue that the shape interpretation of speech-accompanying iconic gestures is dependent on its co-occurrent speech. Since there is no prototypical solution for modeling such a dependence, we offer an approach to compute a gesture's meaning as a function of its speech context.

1 Introduction

Speakers often convey multi-modal content by pointing at things or shaping their contours while talking. The semantics of the verbal part is intertwined not only with the communicative situation and the agent's informational situation, but also with the semantics of the non-verbal part. So, one information providing system (gesture) depends on another one (language) for its interpretation. In gesture research, there are at least three claims about how a gesture's interpretation depends on its accompanying speech context: (i) The *classification* of gestures is speech-dependent (see, e.g., (McNeill, 1992; Kendon, 2004; Müller, 2010; Fricke, 2014)). Whether a movement by the index finger is interpreted as drawing a line or as indexing an area in gesture space depends on the respective utterances. Such a movement is likely to be interpreted as *indexing* when the speaker says 'There is my ball,' but it is likely to be interpreted as a *drawing* if the speaker utters 'The path continues for ten miles.' (ii) The *individuation* of gestures is speech-dependent. For instance, it depends on the context whether one interprets an iterative movement as one gesture or as several directly subsequent ones (an example by Lascarides and Stone (2009): 403). (iii) Lascarides and Stone argue that an interpretation of a gesture's meaning does not only depend on its shape, but also on its *rhetorical connection* to its speech context (e.g., (Lascarides and Stone, 2009)). We set these three types of dependencies aside here. Instead, we argue that there is another type of dependence: The meaning of gestures with respect to their *shape* interpretation depends on their accompanying speech. In this paper, we present an approach how to model this particular meaning-dependence of iconic gestures.

2 The meaning-dependence of iconic gestures on their co-occurring speech

The iconic gestures we are concerned with are spontaneous movements of hands or fingers that do not have a lexical meaning. Here, we employ McNeill's conception of a stroke and its semantic synchrony with the accompanying speech (McNeill, 1992), but we acknowledge the idealizations involved in these matters (for treatments of asynchronous strokes, see, e.g., (Hahn and Rieser, 2012)). We take for granted that modeling the meaning of gestures *qua* linguistic signs requires a well-founded concept of meaning and benefits from a formal semantics approach.

Humans do not gesticulate geometrical shapes. If one takes a closer look at roundish-looking gestures, one quickly notices that such gestures are mostly if not always spiral. If a speaker iterates such a sloppy gesture, it looks helix-like. Moreoever, gestures that are intended to be angular are often roundish. This sloppiness is presumably due to the physiological features of humans, time limits, etc. Despite this fact it is common to interpret gestures as conveying meanings like round′ or square′. It seems natural to interpret, say, a roundish gesture as an imperfect sign for the

meaning round′. Roundish gestures can be interpreted as *approximating* geometrical shapes like circles. If so, the gesture's speech-*independent* morphological features alone, such as its hand shape, movements, could provide the core of the gesture's meaning. This view has been (implicitly or explicitly) suggested by authors of formal theories of gesture meaning (which range from employing HPSG (e.g., (Johnston, 1998; Lücking, 2013; Alahverdzhieva and Lascarides, 2010)), to LTAG (e.g., (Kopp et al., 2004)), to λ-calculus (e.g., (Rieser, 2004)), to Montague grammar (e.g., (Giorgolo, 2010)), to SDRT (e.g., (Lascarides and Stone, 2009))[1], and to TTR (e.g., (Lücking, forthcoming)). One might argue for such an approach by suggesting that humans abstract away from the sloppiness while interpreting gestures, since most if not all gestures are sloppy. Sloppiness itself need not pose a problem (apart from the problem of exact depiction). Nonetheless, we found that the sloppiness is the reason for a specific speech-dependence of gesture meaning. In what follows, we argue that the interpretation of a gesture's *shape* is dependent on the meaning of its accompanying speech. Only interpreted in particular contexts are roundish gestures interpreted as meaning round′ rather than angular′.

First, gestures that share all relevant morphological features (i.e., that are of the same type) can be interpreted differently given different speech contexts. If a helix-gesture accompanies an utterance like 'The window is round' it is likely to be taken as meaning circular′ or round′. If it accompanies 'The townhall features a staircase' it is likely to be interpreted as meaning spiral′. Depending on the standard of precision at stake, a roundish gesture might be interpreted as conveying round′ when accompanied by 'ball', but as conveying angular′ when accompanied by 'box'. Such an ambiguity is also found when the sloppiness of the gesture is extreme. Take a look at the examples given in Fig. 1. In Fig. 1a the speaker is uttering 'But not round spiral staircases, but so eh. If the house is rectangular, can the stairs outside be [truncation].' (English translation, gesture stroke underlined) The emphasis on 'rectangular' and the overlapping stroke together with other parts of the dialogue suggests that the speaker employs the gesture to illustrate the shape of the house. Of course, it is also plausible to interpret her gesture as modeling the house, but that seems dispreferred because of the stroke overlap and the content of the overlapping speech. Interestingly, the same speaker uses a similar gesture also in the following speech context: 'And it stands on such a round base?' (see Fig. 1b) Here, it is again plausible that the gesture illustrates a shape. But this time it seems to illustrate roundness. So, we encounter very similar gestures with quite different meanings due to different speech contexts. Our corpus provides more of these examples. The general observation is that one type of gesture (individuated via a similar gesture annotation) can have different gesture meanings when accompanying different utterance segments:

(a) 'rectangular house' (b) 'round base'

Figure 1: Similar gesture morphology, but different meaning

(I) One type of gesture accompanying different utterance segments has different meanings as value.

Second, gestures with a significantly different gesture morphology can represent the same meaning. For instance, different gestures can convey the meaning rectangular′ if they relate to the same utterance segment, etc. Take as examples the ones shown in Fig. 2. In Fig. 2a the speaker utters 'It is just a rectangular building.' Compare this displaying of rectangular′ with Fig. 2b which is identical to Fig. 1a. Although the gestures display some similarity, they are clearly different. Nonetheless, they both seem to mean rectangular′ or angular′. Here, the general observation is that different types of gesture accompanying the same or semantically similar utterance segments can select the same gesture meaning as value:

(II) Different types of gestures accompanying the same utterance segment have one and the same meaning.

(I) and (II) support the idea that the meaning of an iconic gesture is determined to a significant extent by the meaning of its accompanying speech.[2]

[1] Lascarides and Stone employ annotations featuring geometrical shapes, such as circles and cylinders, for their underspecified gesture meanings (e.g., (Lascarides and Stone, 2009): 402, 407, 430, 436).

[2] Our examples feature single words, but our account is

(a) 'rectangular building'

(b) 'rectangular house'

Figure 2: Different gesture morphology, but same meaning

3 Modeling the dependence

This dependence of gesture meaning on speech meaning has not been modeled. The gesture theories mentioned above could only cope with it by substantially underspecifying the gesture's meaning. This would allow the meaning of, say, a spiral gesture to be compatible with utterance segments with conflicting meanings, such as 'round' and 'rectangular'. But this would render the gesture's meaning too weak. It would not allow for recognizing the gesture's contribution to the communicated content and it would not fit the intuition that iconic gestures have a rich meaning on their own. There is also no prototypical solution to be found in other formal semantics: Formal semantics travels the inverse route, so to speak, modeling the context dependence of speech, whereas we model a dependence on speech as context.

A new model of the meaning of iconic gestures should meet at least the following desiderata: (a) The meaning of a gesture is determined to a significant extent by the meaning of the accompanying speech. A similar gesture morphology is *not* sufficient for a similar/identical meaning and a different gesture morphology is *not* sufficient for a different meaning. (b) Nonetheless, its morphology is not irrelevant for determining a gesture's meaning. Not just any gesture can have the meaning round′, for instance, a clearly articulated angular gesture cannot. So, a gesture's meaning is not completely determined by speech. Moreover, gesture content can *contradict* speech meaning. Our corpus has one remarkable instance in which a 'cup-upwards-word' is accompanied by a 'cup-down-wards' gesture.

From a formal point of view, (II) does not present new obstacles over and above those encountered in the context of observation (I). A roundish gesture accompanying, say, 'clock' or 'window' could either be drawn with one index-finger or shaped or modeled with both hands. According to our annotation practices, these would be different gestures, in part due to the different handshapes used. In addition, more subtle differences in terms of gesture morphology could arise. According to the account presented here, the different gestures might all yield ⟦round⟧ if combined with ⟦clock⟧ or ⟦window⟧.[3] Arguments supporting that would have to be given for (I), too.

For (I) our account has to specify the speech-dependent meaning of the gesture. Here is an outline of our approach: The gesture meaning is a *function* of the gesture's initial (topological) meaning based on its morphology and the speech context. The gesture's morphology is described by attribute-value pairs (AVMs) concerning hand shape, movements, etc. One computes the *initial meaning* of the gesture mapping the AVMs onto a logical formula. The *final gesture meaning* is a function of the initial meaning and the speech context. Then, speech meaning and final gesture meaning can be combined to gain a multi-modal proposition (see Fig. 3).

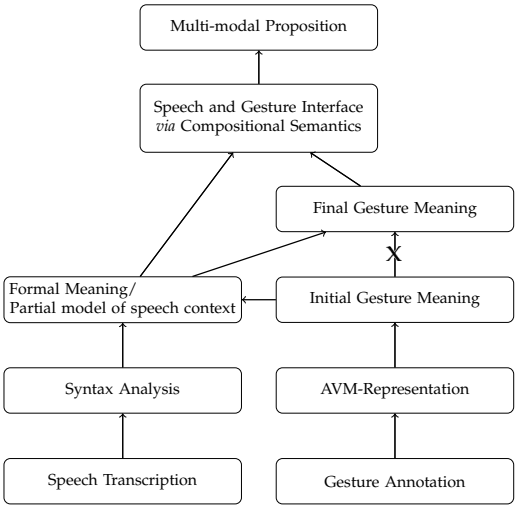


Figure 3: Methodology of our approach

Such an approach is best pursued using a dynamic semantics, because we need a device that is able to model the evolution of the interpretation of gesture processes and speech processes as well as their interaction. The interaction handles compositionality of non-speech and speech meanings. No known static semantics can fulfill such desiderata. We use the ψ-calculus, a recent extension of Milner's π-calculus (see, (Milner, 1999; Johansson, 2010; Rieser, 2015)). ψ has concurrent chan-

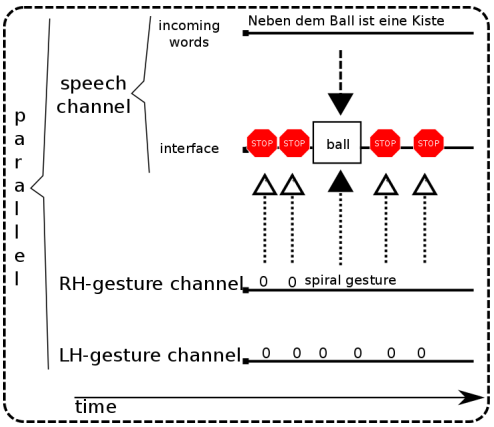
not, in principle, restricted to gesture-word relations.

[3] '⟦A⟧' denotes A's extension; 'A'' the whole meaning.

nels to transmit and process information specified as data structures. Channels are the input-output-devices known from concurrent programming. We represent channels as ψ-operators. They can transport any logic information, such as expressions of a typed λ-calculus and their partial models.

Implementing our approach roughly works as follows: The initial semantics of the gesture formulated in λ-terms is passed onto a channel containing the gesture's speech context. The speech context may modify the gesture meaning in various ways (see, e.g., (I)). Assume that the gesture's initial meaning is spiral$'$, its speech context ball$'$. Roughly, ball$'$ is sent to spiral$'$ which changes it to round$'$ and finally uses it as a modifying information. So, transported and modified meanings are treated in the end as fixed points. In Fig. 4 you can see the basic idea illustrated. The example utterance is 'Neben dem Ball ist eine Kiste.' (Engl.: 'Next to the ball there is a box.') As shown in Fig. 4a, the idea is that a spiral gesture in the context of objects like ⟦ball⟧ and other roundish things designates round$'$ (observe the use of meta-language and object language expressions here which is vital) and $\perp$ (undefined) else. So, the multi-modal meaning of 'ball' + spiral gesture is ball$'(x) \wedge$ round$'(x)$. More specifically, if the partial model input '⟦...⟧' to (2), instantiating bae, yields $z \in \{$⟦circle⟧, ⟦clock-face⟧, ⟦mirror⟧, ⟦sign⟧, ⟦ball⟧, ⟦cup-bottom⟧, ...$\}$ and the projection of spiral$'$, f(spiral$'$), approximates circle$'$ in context c to degree $r \geq$ the threshold in c then round$'$ is substituted for ro, [round$'/ro$], and output on ch$_2$; else $\perp$ is substituted for ro, $[\perp/ro]$, and is output on ch$_2$. The $z \in$ clause and the threshold shall guarantee that not just any gesture can mean round$'$. Gestures accompanying a phrase whose extension is not an element of the set (say, 'square'), as well as gestures that do not approximate a circle to the context-sensitive threshold cannot mean round$'$. The threshold can be determined algorithmically through a simulation device as shown in Pfeiffer et al. (2013) for two-dimensional cases. For three-dimensional cases we still rely on intuition.

This account is not an underspecification account of gesture meaning. We suggest a change of the initial meaning gained from the described morphology. It is triggered by the meaning of the accompanying speech, given that restrictions like the satisfaction of an approximation function hold.



(a) Basic intuition: contact point of spiral gesture and word 'ball'. The spiral gesture + 'ball' yields round$'$, according to (b).

$$\underline{ch}_1\ bae\ \overline{ch}_2\ ro\ <\ \lambda z \exists f \exists c \exists r \exists \text{thr}_c\ (\text{spiral}' \wedge \text{approximates}(f(\text{spiral}'), c, x) = r \wedge r \geq \text{thr}_c \wedge \text{circle}'(x) \wedge \text{context}(c) \wedge z \in \{[\![\text{circle}]\!], [\![\text{clock-face}]\!], [\![\text{mirror}]\!], [\![\text{sign}]\!], [\![\text{ball}]\!], [\![\text{cup-bottom}]\!], \ldots\} \rightarrow [\text{round}'/ro][else][\perp/ro] > (bae)$$

(b) If-else rule for interpreting a spiral gesture in the context of, say, ⟦ball⟧ as round$'$

Figure 4: Modeling with the λ-ψ-calculus.

4 Conclusion and further research

We argued that gestures have a speech-dependent meaning and proposed to model their meanings as a function of the gesture's initial meaning and the speech context employing the ψ-calculus. On account of this, gestures with the same morphology can have even conflicting meanings if they appear in different speech contexts, e.g., we can assign meanings like rectangular$'$ vs. circular$'$ to similar gestures. For future research we aim at integrating the speech context's influence on the gesture classification and individuation as well as the role of rhetorical relations, and at expanding our model for analyzing more complex gestures.

5 Acknowledgements

We are grateful to three reviewers for their critical comments. We tried to accommodate most, but some suggestions, such as the generalization of the model to discourse data (see, however, (Rieser, 2017)) or to the speaker's perspective during the production of co-speech gestures or discussing whether speech disambiguates gesture meaning rather than changing its initial meaning, have to be tackled on another occasion.

References

Katya Alahverdzhieva and Alex Lascarides. 2010. Analysing language and co-verbal gesture in constraint-based grammars. In Stefan Müller, editor, *Proceedings of the 17th International Conference on Head-Driven Phase Structure Grammar (HPSG)*, pages 5–25, Paris.

Ellen Fricke. 2014. Between reference and meaning: Object-related and interpretant-related gestures in face-to-face interaction. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem, editors, *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, pages 1788–1802. De Gruyter Mouton.

Gianluca Giorgolo. 2010. *Space and Time in Our Hands*. UIL-OTS, Universiteit Utrecht.

Florian Hahn and Hannes Rieser. 2012. Non-compositional Gestures. In *International Workshop on Formal and Computational Approaches to Multimodal Communication held under the auspices of ESSLLI 2012*, Opole.

Magnus Johansson. 2010. Psi-calculi: a framework for mobile process calculi.

Michael Johnston. 1998. Unification-based multimodal parsing. In *Proceedings of the 36th Annual Meeting on Association for Computational Linguistics*, volume I, pages 624–630, Montreal, Quebec. ACL.

Adam Kendon. 2004. *Gesture – Visible Action as Utterance*. Cambridge University Press, Cambridge, NY, fourth edition.

Stefan Kopp, Paul Tepper, and Justine Cassell. 2004. Towards integrated microplanning of language and iconic gesture for multimodal output. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 97–104, New York, NY. ACM.

Alex Lascarides and Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics*, 26(4):393–449.

Andy Lücking. 2013. *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. De Gruyter Mouton, Germany.

Andy Lücking. forthcoming. Modeling co-verbal gesture perception in type theory with records. In *Proceedings of the 1st International Workshop on AI Aspects of Reasoning, Information, and Memory*, AIRIM'16. Federated Conference on Computer Science and Information Systems.

David McNeill. 1992. *Hand and Mind. What Gestures Reveal About Thought*. The University of Chicago Press.

Robin Milner. 1999. *Communicating and mobile systems: the π-calculus*. Cambridge University Press.

Cornelia Müller. 2010. Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. *Sprache und Gestik. Sonderheft der Zeitschrift Sprache und Literatur*, 41(1):37–68.

Thies Pfeiffer, Florian Hofmann, Florian Hahn, Hannes Rieser, and Insa Röpke. 2013. Gesture semantics reconstruction based on motion capturing and complex event processing: a circular shape example. In *Proceedings of the SIGDIAL 2013 Conference*, pages 270–297, Metz, France. Association for Computational Linguistics.

Hannes Rieser. 2004. Pointing in dialogue. In *Proceedings of Catalog 04. The 8th Workshop on the Semantics and Pragmatics of Dialogue*, pages 93–101, Barcelona, July.

Hannes Rieser. 2015. When hands talk to mouth. gesture and speech as autonomous communicating processes. In *Proceedings of the 19th workshop on the semantics and pragmatics of dialogue*, pages 122–130.

Hannes Rieser. 2017. A Process Algebra Account of Speech-gesture Interaction. Revised and Extended Version. In *International Workshop on Formal Approaches to the Dynamics of Linguistic Interaction held under the auspices of ESSLLI 2017*, Toulouse.