

Focus Marking via Gestures

Cornelia Ebert
University of Stuttgart

Stefan Evert
University of Osnabrück

Katharina Wilmes
University of Amsterdam

Our talk contributes to the recent investigations of speech-accompanying gestures under a formal semantic view (Lascarides and Stone 2009, Rieser 2008). We will show that gestures can serve to disambiguate a sentence with respect to the possible focus projections by a statistical evaluation of data gained from a corpus annotated with gestures and information structure. The language under investigation is German. We argue that a sentence that, in isolation, is ambiguous concerning the extension of its focus domain is disambiguated via speech-accompanying gestures. Gesture thus is a means to mark information structure next to intonation and word order.

It is widely known that gestures are temporally aligned with the speech signal, i.e. the stroke of a gesture falls together with the main accent of the gesture-accompanying sentence (McNeill 1992 among many others). The relationship of complete gestures and foci, however, has not been investigated systematically yet. In our talk, we fill this gap by showing that the focus of a linguistic expression projects depending on the onset of the accompanying gesture. Gesture thus serves as a means to mark focus domains.

Consider the following example for illustration (main accent indicated by capital letters):

(1) I ate baNAnas.

The sentence in (1) with the given intonation pattern can be read as an answer to the two questions in (2), each inducing a different focus-background structure (cf. Selkirk 1982 for the underlying focus projection rules).

(2) a. What did you do?

b. What did you eat?

(2a) is a VP-focus invoking question, while (2b) requires narrow focus on the direct complement. Concerning (1), we arrive at the focus patterns exemplified in (3) as answers to (2a-b), respectively.

(3) a. I [ate baNAnas]_F.

b. I ate [baNAnas]_F.

Our initial hypothesis was that the onset of a speech-accompanying gesture (independent of the type of gesture – be it a beat, an iconic gesture or another kind of gesture) serves to indicate the left border of a focus phrase and can thus serve to disambiguate an information-structural ambiguity in a sentence towards a certain focus-background pattern. Simplifying matters a little, we expected the pattern in (4). (_G marks the hypothesized onset of the speech-accompanying gesture.)

(4) a. I |_G[ate baNAnas]_F.

b. I ate |_G[baNAnas]_F.

As can be seen, although (1) is ambiguous with respect to the underlying information structure, /_G disambiguates the sentence towards one of the focus-background patterns in (3).

In order to test our hypothesis, we looked at the temporal relationship between the onset of a gesture and the beginning of its associated focus. We further annotated the multimodal Bielefeld Speech-And-Gesture-Alignment (SAGA) corpus, which was already annotated with gestures, with focus features and nuclear accents using the software tools Elan and Praat,

respectively. A subsequent statistical analysis confirmed our hypothesis that focus and gesture align. (We could also confirm the result of previous studies that the stroke of a speech-accompanying gesture and nuclear accent fall together.) However, somewhat surprisingly we also found that there is a systematic shift – in general gestures start about 0.31 seconds earlier than foci, with a standard deviation of 0.39 seconds. That is, there is a certain time lag between the onsets of gesture and focus, but in most cases they are within less than one second of each other and can thus be considered to be aligned. A one-sample t-test showed that the time lag is significant ($t = 13.0792$, $df = 261$, $p < .001$; H_0 : mean time lag = 0). The corresponding 95%-confidence interval places the true mean time lag between gesture and focus in the range from 0.271s to 0.367s. (We verified that the distribution of time differences is sufficiently close to a Gaussian for the t-test to be applicable.) The significant time shift we found might have its roots in the fact that it allows the hearer to draw more attention to the phrase in focus, as its occurrence is made predictable by a preceding gesture.

Our work makes a contribution to the general research question of how speech-accompanying gestures have to be dealt with formally. It is widely accepted that speech and gestures work together to convey one single thought (see e.g. McNeill 1992, Kendon 2004) and that the semantic content of these gestures is intertwined with the semantic content of the speech signal. In the literature, the semantic contribution of gestures is usually illustrated with referential gestures or speech accompanying iconic gestures, which express semantic content. Speech accompanying iconic gestures can be co-expressive (displaying the same semantic content as the speech signal), but also complementary (expressing additional information). Often iconic gestures are only interpretable in combination with the speech symbol (Kopp et al 2004, Lascarides & Stone 2009), and this strong interaction has often been interpreted in such a way that information from the two channels – speech and gesture – is mapped to one single logic representation (e.g. in Rieser 2008, Kopp et al 2004). Independent of the formal details of the implementation, we support the view of (Lascarides and Stone 2009) that the contribution that gestures make to the semantics of an utterance can only be dealt with correctly in a formal semantic model that integrates the usual semantics of the involved linguistic expressions and the semantic contribution of their accompanying gestures. While (Lascarides and Stone 2009) show that only an integrated model can capture the scope relations of an iconic gesture and a linguistically given negation and that only such a model can explain how iconic gestures specify certain rhetoric relations that are underspecified in the linguistic material, (Engle 2000) shows that iconic gestures can serve to disambiguate a sentence towards a distributive reading. Our work adds to this line of research in showing that gestures mark focus domains and hence have to be treated in an integrated model. Additionally, our work shows that also simple beats can make a contribution to the semantics: focus projection is generally restricted by the onset of a speech-accompanying gesture, be it iconic, referential, or a simple beat.

- Engle, Randi A. (2000): *Toward a Theory of Multimodal Communication: Combining Speech, Gestures, Diagrams and Demonstrations in Structural Explanations*. Ph.D. thesis.
- Kendon, Adam (2004): *Gesture. Visible Action as Utterance*. Cambridge University Press.
- Kopp, Stefan, P. Tepper & J. Cassell (2004): Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. *Proceedings of the International Conference on Multimodal Interfaces (ICMI'04)*, S. 97-104.
- Lascarides, Alex & M. Stone (2009): A Formal Semantic Analysis of Gesture, *Journal of Semantics* 26(4), 393-449.
- McNeill, David (1992): *Hand and Mind. What Gestures Reveal About Thought*.
- Rieser, Hannes (2008): Aligned Iconic Gesture in Different Strata of Multi-Modal Route-description dialogue. In: *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue*.
- Selkirk, Elisabeth (1984): *Phonology and Syntax: The Relation between Sound and Structure*.