Putting the notion of context-free gesture meaning to test

Florian Hahn, Insa Lawler and Hannes Rieser*

Collaborative Research Center "Alignment in Communication" (CRC 673) Bielefeld University, Germany

{fhahn2|ilawler|hannes.rieser}@uni-bielefeld.de

Introduction

Face-to-face communication is often accompanied by gestures. Intuitively, gestures have meaning and contribute to the communicated content. But what is a gesture's meaning and how is it determined?

One common view is as follows: Gestures are signs which have a *standalone* context free meaning (e.g., Johnston et al. (1997), Kopp et al. (2004), Rieser (2004)). We doubt that this view is accurate and motivate that a different one should be taken: the meaning of a gesture is determined to a significant extent by the meaning of the accompanying speech.

In the following, we explicate a version of the simple view and adduce reasons why to believe that it is inaccurate. We then outline an ongoing study which aims to investigate whether a different view on modelling a gesture's meaning is empirically supported.

The simple view

In our previous research (e.g., Hahn et al. (2014)), we advocated a version of the simple view. A gesture's meaning is gained mainly independently of the meaning of its accompanying speech. The meaning of a gesture is determined by considering its stroke, i.e., its meaningful part. This part is characterized by gesture morphological predicates and values, such as Hand_Shape, Wrist_Position or Movement_Direction. In addition, we need one further predicate: the Representation_Technique. The Representation_Technique determines relevant morphological predicates and respective values for the translation of the predicate-value pairs into a logical formula. The Representation_Technique "drawing" highlights movements of the index finger. In order to calculate a gesture's meaning, the combinations of its predicate-value pairs are translated into a logical formula. For instance, the predicate-value pairs `[Representation_Technique, Drawing]` and `[Path_of_Movement, Line]` can be translated together into the formula $\exists x(trajectory(x) \wedge straight(x))$. So, the standalone meaning of this predicate-value combination is to represent a straight trajectory. This meaning can be combined with the meaning of the co-occurring speech in order to yield a multi-modal meaning.

Given such a methodology, the only part of gesture meaning which depends on the accompanying speech is its Representation_Technique: whether a gesture is interpreted as a static (e.g., indexing) or dynamic one (e.g., shaping), depends crucially on the recipient's interpretation of the gesture considering the speech context, e.g., whether a direction or place is indicated or a surface shaped. Since the gesture's stroke is partly determined by its Representation_Technique, the speech context has a non-negligible influence on the determination of the gesture's meaning. Nevertheless, the morphology of the gesture stroke is determined in a context free way. A linear movement will always result in the predicate-value

*Authors in alphabetical order.

[Path_of_Movement, Line].

This picture seems inaccurate for the following reasons: if one takes a closer look at the classification of a gesture's *shape* our corpus data suggest that it is dependent on the co-occurring speech. Humans never gesticulate perfect geometrical shapes. Gestures which are intended to be circular are mostly spiral gestures. A spiral gesture, however, can be interpreted differently. If a spiral gesture is accompanied by the utterance "The window is round" it is more likely to be interpreted as CIRCULAR, whereas the *same* gesture accompanied by "The staircase looks nice" would render a gesture meaning like SPIRAL. Thus, the meaning of the same gesture morphological predicates varies with the speech context. In order to clarify whether these observations are accurate, we conduct an empirical study the basic idea of which is described in the following. If the empirical findings support our hypothesis, we aim to construct a new model for determining a gesture's meaning.

Examining the new view

The aim of our study is to investigate whether speech meaning influences the interpretation of a gesture's meaning beyond guiding the interpretation of the Representation_Technique used. In the study, we focus on the (potential) influence of the verbal meaning on the classification of the *shape* of iconic gestures, specifically on circular and rectangular drawing gestures.

Our basic method is as follows: we construct multi-modal utterances by combining the video recording of the torso of a speaker with different records of her head, in which she utters different sentences. We make use of gestures which draw trajectories that resemble simple geometrical shapes like circles and rectangles. We also make use of gestures that could be considered to approximate both a square and a circle (e.g., squares with very rounded down corners). The utterances either contain NPs which denote objects which have a prototypical circular shape (e.g., "ball") or a prototypical rectangular shape (e.g., "box"), or which can have various shapes ("table"). For the combinations of torso and head we then have two variants. In one variant, the gesture is the constant factor, i.e., one and the same gesture is combined with different sentences. In the other variant, the speech context is held fixed, i.e., one and the same utterance is combined with different gestures.

If it turns out that the participants indeed interpret one and the same gesture differently in different speech contexts, our new view on gesture meaning is supported.

References

Hahn, F., Lawler, I., and Rieser, H. (2014). Speech-gesture-interface constructions for gestures accompanying german verb phrases. *Presented at the 6th Conference of the International Society for Gesture Studies (ISGS 2014), University of California, San Diego.*

Johnston, M., Cohen, P. R., McGee, D., Oviatt, S. L., Pittman, J. A., and Smith, I. (1997). Unification-based multimodal integration. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, pages 281–288, Madrid. ACL.

Kopp, S., Tepper, P., and Cassell, J. (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.

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