Abduction and parameterised semantic composition in speech-gesture integration

Udo Klein, Hannes Rieser, Florian Hahn and Insa Lawler
Collaborative Research Center
"Alignment in Communication" (CRC 673)
Bielefeld University, Germany

Abstract

An important feature of speech-gesture integration is that speech and gesture content influence each other in interpretation. To analyse this, we sketch an approach based on three assumptions: (i) the interpreter infers by abduction an explanation for why a particular gesture is synchronized with a particular utterance (part), (ii) semantic composition amounts to the identification of free variables (called parameters), and (iii) abductive inferences determine which parameters to identify during the semantic composition of speech and gesture content.

1 Introduction

Face-to-face communication is often accompanied by gestures: Speakers point at things or shape their contours. Foundational questions arise: What is a gesture's meaning and how is it determined? And, given that speech and gesture meaning interact, how can they be fused? The issue of speech-gesture integration (SGI) has been studied in various paradigms such as Montague Grammar, HPSG and theories of Discourse and Dialogue; it is also the focus here. We will demonstrate a methodology for integrating verb phrases with accompanying gestures based on parameterised semantic composition.

Our work is based on data from a systematically annotated corpus, the Bielefeld-Speech-and-Gesture-Alignment-corpus (SaGA; (Lücking et al., 2013)), which consists of 25 dialogues of dyads engaged in route descriptions. Consider the following example (cf. Fig. 1). The speaker in Fig. 1 describes how to walk through a park passing a pond. While uttering *Gehst quasi drei Viertel um den Teich herum* (Engl.: '(You) roughly walk three quarters around the pond (around)'),

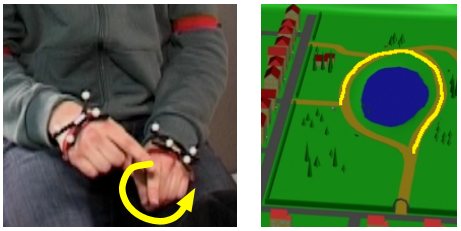

Figure 1: Gesture (left) depicting the agent's trajectory around a pond (right)

a round shape is depicted in overlap with the expression *drei Viertel um den Teich herum*. Importantly, this expression specifies that the agent's trajectory is three quarters around the pond, but it does not specify the actual shape of the trajectory (the shape of the pond does not necessarily determine the shape of the agent's trajectory around the pond). As a result of being synchronized with this expression, the gesture can be interpreted as specifying that the shape of the agent's trajectory around the pond is circular. To analyse this, we propose (i) that the interpreter infers by abduction an explanation for why the gesture is synchronized with this utterance part, (ii) semantic composition amounts to the identification of free variables (called parameters), and (iii) the abductive inference enriches semantic composition of speech and gesture content by determining which parameters to identify. In our example the inferred explanation for the synchronicity of gesture and utterance is that the finger trajectory approximates the shape of the agent's trajectory around the pond.

2 Motivating parameterised semantics for SGI

Previously, we have developed a general methodology for SGI, abstracting from speech acts. We worked out a λ-calculus based solution in which speech meaning is type-lifted to a function which takes gesture meaning as an argument and yields

the integrated meaning (cf. Röpke et al., 2013). Here, we present an alternative approach in order to explicitly model the way in which abductive inferences enrich the semantic composition of speech and gesture content (cf. Hobbs (2008) for an overview of abduction in natural language understanding). We propose that the basic principle of semantic composition is conjunction (cf. Pietroski, 2005) relative to (i) a coordination scheme (cf. Fine, 2007) which specifies which free variables in the conjuncts are to be identified, and (ii) a systematic renaming of the remaining free variables in order to avoid accidental identification (cf. Kracht, 2013). To illustrate, the composition of the two formulas $P(x4, x5)$ and $Q(x5, x6)$ relative to the coordination scheme $\{\langle x4, x6\rangle\}$ results in the formula $P(x40, x50) \wedge Q(x51, x61) \wedge x40 = x61$. The free variables in the left and right conjuncts have been suffixed by a 0 and 1, respectively, in order to avoid the accidental identification of the two $x5$ occurrences. The coordination scheme indicates which free variables get identified. An important consequence of using parameterised semantic composition in SGI is that speech and gesture content can be used to instantiate rules used in abductive inferences, and thus to determine the parameter(s) of the utterance content that the gesture content specifies. Moreover, speech and gesture content can be combined without having to change the combinatory potential (and thus the logical type) of utterance content.

3 Analysis

Applying this theory to our example, the composition of $[\![$ *drei Viertel* $]\!]$ and $[\![$ *um den Teich herum* $]\!]$ conjoins the two formulas and identifies the degree parameters by adding the equation $d0 = d1$:

$$\frac{drei\ Viertel}{\underline{d} = 0.75} \bullet_{\{\langle \underline{d}, \underline{\underline{d}}\rangle\}} \frac{um\ \ldots\ herum}{\begin{array}{l}\mathsf{mover}(e) = x \wedge\\ \mathsf{trajectory}(x, e) = t \wedge\\ \mathsf{around}(t, r, d) \wedge\\ r = \imath x.\mathsf{pond}(x) \wedge\\ \underline{\underline{d}} \geq 0.5\end{array}} =$$

$$\frac{drei\ Viertel\ \ldots\ herum}{\begin{array}{l}d0 = 0.75 \wedge\\ \mathsf{mover}(e1) = x1 \wedge\\ \mathsf{trajectory}(x1, e1) = t1 \wedge\\ \mathsf{around}(t1, r1, d1) \wedge\\ r1 = \imath x.\mathsf{pond}(x) \wedge\\ d1 \geq 0.5 \wedge\\ \underline{d0} = \underline{d1}\end{array}}$$

($x1$ is the entity moving in $e1$, $x1$'s trajectory in $e1$ is $t1$, $t1$ circumscribes some pond $r1$ to a degree $d1 \geq 0.5$, and $d1 = 0.75$.)

The semantic integration of speech and gesture is based on an abductive inference involving the following gesture interpretation rule (GIR):

GIR If parameter p of gesture content $[\![G]\!]$ approximates some parameter p' of utterance content $[\![U]\!]$, then G is synchronized with U.

The most plausible instantiation of **GIR** in the utterance context is that the parameter g of $[\![G_1]\!]$ representing the finger trajectory approximates the parameter t of $[\![U_1]\!]$ representing the mover's trajectory around the pond. Since G_1 is synchronized with U_1, the interpreter can infer by abduction that indeed the parameter g of $[\![G_1]\!]$ (the finger trajectory) approximates the parameter t of $[\![U_1]\!]$ representing the mover's trajectory around the pond. This inference enriches the semantic composition of gesture and utterance by adding the formula $\mathsf{approx}(g, t)$ to the gesture content, and by specifying the coordination scheme for composition, namely that the trajectory $\underline{t}$ which g approximates is to be identified with the trajectory $\underline{\underline{t}}$ of the mover x in e:

$$\frac{G}{\begin{array}{l}\mathsf{circular.traj}(g) \wedge\\ \mathsf{approx}(g, \underline{t})\end{array}} \bullet_{\{\langle \underline{t}, \underline{\underline{t}}\rangle\}} \frac{drei\ \ldots\ herum}{\begin{array}{l}\mathsf{mover}(e) = x \wedge\\ \mathsf{traj}(x, e) = \underline{\underline{t}} \wedge\\ \mathsf{around}(t, r, d) \wedge\\ r = \imath x.\mathsf{pond}(x) \wedge\\ d = 0.75\end{array}} =$$

$$\frac{G + drei\ \ldots\ herum}{\begin{array}{l}\mathsf{circular.traj}(g0) \wedge\\ \mathsf{approx}(g0, t0) \wedge\\ \mathsf{mover}(e1) = x1 \wedge\\ \mathsf{traj}(x1, e1) = t1 \wedge\\ \mathsf{around}(t1, r1, d1) \wedge\\ r1 = \imath x.\mathsf{pond}(x) \wedge\\ d1 = 0.75 \wedge\\ \underline{t0} = \underline{t1}\end{array}}$$

The resulting multimodal representation thus expresses that the mover's trajectory around the pond is a circular one.

To conclude, we propose a novel approach to speech-gesture integration, in which the gesture interpretation is determined by context-dependent abductive inferences and gets integrated with the utterance denotation by parameterised semantic composition. In future work, we intend to compare this approach with our λ-calculus based approach, focusing in particular on how these approaches explain the fact that speech and gesture interpretation mutually influence each other.

Acknowledgments

This research is supported by the German Research Foundation (DFG) in the Collaborative Research Center 673 “Alignment in Communication”.

References

Fine, K. (2007). Semantic Relationism. Blackwell, Oxford.

Hobbs, J. R. (2008). Abduction in Natural Language Understanding. In Horn, L. and Ward, G., editors, The Handbook of Pragmatics, pages 724–741. Blackwell Publishing Ltd.

Kracht, M. (2013). Agreement Morphology, Argument Structure and Syntax. Unpublished manuscript.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., and Rieser, H. (2013). Data-based Analysis of Speech and Gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its Applications. Journal on Multimodal User Interfaces, Vol. 7(1-2), pages 5–18.

Pietroski, P. (2005). Events and Semantic Architecture. Oxford University Press.

Röpke, I., Hahn, F., and Rieser, H. (2013). Interface constructions for gestures accompanying verb phrases. In Proceedings of 35th Annual Conference of the German Linguistic Society (DGfS), pages 295–296.