

# Multimodal clarification potential

Jonathan Ginzburg & Andy Lücking  
Université Paris Cité & Goethe University Frankfurt

Keywords: non-verbal repair, laughter, gesture, clarification request, multimodal interaction

Non-verbal other repair is rare by comparison with verbal repair and, in part, due to this understudied (though see Kimbara, 2006; Healey et al., 2015; Hough et al., 2015; Mazzocconi et al., 2018a; Ginzburg and Lücking, 2021). Thus, while the frequency of verbal clarification questions ranges from 3–7 % in adult-adult corpora (Ginzburg, 2012), the frequency of such questions with respect to laughter is less than 0.04 % of all the laughs produced (Mazzocconi et al., 2018b) (we are not aware of corresponding estimates for manual gestures). Nonetheless, it can and does occur in manual gesture, laughter, motor skill coaching (and a variety of other cases):



(1) a. A: I think this should be [  ] staircases. B:  [spiral gesture]? A: A spiral staircase. (Based on example from SaGA Corpus, Lücking et al., 2013)

b. A: I hear you're busy (laughter) [= little giggle]. B: (laughter)? (= low arousal laughter with rising contour). (Attested example from Ginzburg and Lücking, 2021)

c. Coach: feet further apart [moves feet]. Coachee: [moves feet and looks at coach with raised eyebrows]. Coach: exactly a bit further apart. (Based on Hough et al., 2015)

Our aim in this paper is to explicate the potential for multimodal clarification and posit constructions that enable its formal description. We build on an account of verbal clarification interaction in Purver (2006); Ginzburg (2012) and of multimodal representation in Lücking and Ginzburg (2023).

Our basic claims will be the following:

- As far as **confirmation clarification questions** (*Did you mean ...*) go: a gestural source allows for verbal clarification questions and equally a gestural source allows for verbal clarification questions.
- For **intended content clarification questions** (*What do you mean ...*), formal parallelism is required: a verbal (gestural) requires a verbal (gestural) clarification question (though accompanying gestural/verbal material is possible)
- Gestures that constitute demonstrations (= main content, in scope of *like this*, instructions, ...) or contribute to compositional content give rise to clarification, whereas those that merely accompany speech (e.g., beats) do not.

We will show that dialogical rules posited for speech are sufficient to explicate the potential for gestural clarification questions, in particular the lack of clarification potential for beats. Similarly, assuming multimodal semantic composition rules (Lücking and Ginzburg, 2023), existing rules for generating confirmation clarification questions scale up directly to enable gestural confirmation questions. However, generating *intended content clarification questions* requires a novel multimodal notion of formal parallelism, which we formulate.

**Acknowledgments** Support by the *Deutsche Forschungsgemeinschaft* (DFG), grant number 502018965 and the French *Investissements d'Avenir-Labex EFL* program (ANR-10-LABX-00), is gratefully acknowledged.

## References

Ginzburg, Jonathan. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Ginzburg, Jonathan and Andy Lücking. 2021. Requesting clarifications with speech and gestures. In *Proceedings of the 1st Workshop on Multimodal Semantic Representations (MMSR)*, pages 21–31.

Healey, Patrick George, Nicola Plant, Christine Howes, and Mary Lavelle. 2015. When words fail: collaborative gestures during clarification dialogues. In *2015 AAAI Spring Symposium Series: Turn-Taking and Coordination in Human-Machine Interaction*.

Hough, Julian, Iwan de Kok, David Schlangen, and Stefan Kopp. 2015. Timing and grounding in motor skill coaching interaction: Consequences for the information state. In *Proceedings of the 19th SemDial Workshop on the Semantics and Pragmatics of Dialogue (goDIAL)*.

Kimbara, Irene. 2006. On gestural mimicry. *Gesture* 6(1):39–61.

Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based analysis of speech and gesture: The Bielefeld speech and gesture alignment corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces* 7:5–18.

Lücking, Andy and Jonathan Ginzburg. 2023. Leading voices: Dialogue semantics, cognitive science, and the polyphonic structure of multimodal interaction. *Language and Cognition* 15(1):148–172.

Mazzocconi, Chiara, Vladislav Maraev, and Jonathan Ginzburg. 2018a. Laughter repair. In L. Prévot, M. Ochs, and B. Faire, eds., *Proceedings of AixDial, the 22nd Workshop on the Semantics and Pragmatics of Dialogue*, pages 16–25.

Mazzocconi, Chiara, Vladislav Maraev, and Jonathan Ginzburg. 2018b. Laughter repair. In L. Prévot, M. Ochs, and B. Faire, eds., *Proceedings of SemDial 2018 (AixDial)*. Heriot Watt University.

Purver, Matthew. 2006. Clarie: Handling clarification requests in a dialogue system. *Research on Language & Computation* 4(2):259–288.