

Making invisible "trouble" visible: Self-repair increases abstraction in dialogue

Gregory Mills, Gisela Redeker

University of Groningen

A central finding in research on dialogue is that interlocutors rapidly converge on referring expressions which become progressively contracted and abstract. This occurs for a wide range of referents, e.g. spatial locations (Garrod and Doherty, 1994; (2016), music (Healey et al., 2007), concepts (Schwartz, 1995), confidence (Fusaroli et al., 2012), and temporal sequences (Mills, 2011). Cumulatively, these findings suggest that processing that occurs in dialogue places important constraints on the semantics of referring expressions.

However, there is currently no consensus about how best to account for how convergence develops. The iterated learning model of Kirby et al (2002) explains convergence as arising out of individual cognitive biases; the interactive alignment model of Pickering and Garrod (2004) favours alignment, while the collaborative model of Clark (1996) emphasizes the role of positive feedback. By contrast, Healey et al., (2007) argues that negative evidence of understanding plays the central role: When interlocutors initiate repair, this allows them to interactively identify, diagnose and resolve differences in interpretation between conversational partners. Addressing these differences accelerates convergence.

To investigate in closer detail how negative evidence might contribute toward convergence, we report a variant of the "maze task" (Pickering and Garrod, 2004). Participants communicate with each other via an experimental chat tool which automatically transforms participants' private turn-revisions into public self-repairs that are made visible to the other participant. For example, if a participant, A types:

A: Now go to the square on the left, next to the big block on top

and then before sending, A revises the turn to:

A: Now go to the square on the left, next to the third column

The chat server automatically detects the revised text and inserts a hesitation marker (e.g. "umm" or "uhhh" immediately preceding the revision). This would yield the following turn, sent to B:

A: Now go to the square on the left next, to the big block on top umm..I meant next to the third column

Two self-repair formats were used:

(1) A: original turn + hesitation marker + reformulated turn

(2) A: original turn + hesitation marker + "I meant" + reformulated turn

Interventions were performed symmetrically on both members of a dyad. No participants detected the experimental manipulation. Examining the transcripts showed that participants who received these transformed turns used more abstract Cartesian location descriptions than participants in a baseline condition. This pattern was already apparent after 5 minutes in the task. Task performance followed a different pattern – initially participants who received these interventions performed worse – completing fewer mazes and requiring more moves to solve each maze. However, by the end of the task, participants who received the interventions performed at the same level as participants in the baseline condition. Crucially, participants who received transformed turns continued to use more abstract descriptions

We argue that this effect is due to the artificial self-repairs having a beneficial effect of amplifying naturally occurring signals of miscommunication: the artificially generated disfluencies and reformulations are used by participants as cues that their partner is having difficulty coordinating on the semantics of referring expressions. Consequently, participants expend more effort to address these putative problems – and once these problems have been identified and resolved, dyads are able to converge quicker on more stable and more abstract referring schemas.

References:

Clark, H. H. (1996). *Using language*. 1996. Cambridge University Press: Cambridge

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms quantifying the benefits of linguistic coordination. *Psychological science*, 0956797612436816.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181-215.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292-304.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31(2), 285-30

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language* (pp. 121-147). Springer London.

Mills, G. J. (2011, September). The emergence of procedural conventions in dialogue. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*. Cognitive Science Society (pp. 210-211).

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, 141, 52-66.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, 4(3), 321-354.