Amplifying signals of misunderstanding improves coordination in dialogue

Gregory Mills
Centre for Language & Cognition Groningen (CLCG)
University of Groningen
Netherlands
g.j.mills@rug.nl

Gisela Redeker
Centre for Language & Cognition Groningen (CLCG)
University of Groningen
Netherlands
g.redeker@rug.nl

Abstract

We report a dialogue task which investigates how the mechanisms of miscommunication contribute toward referential coordination. Participants communicate via a text-based instant messaging tool which is used to identify turns that were edited prior to sending. These turns are transformed by the server into artificial self-corrections, and sent to the participants. The patterns observed in the dialogues show that these interventions have a beneficial effect on referential coordination.

1 Introduction

A central finding in research on dialogue is that interlocutors rapidly converge on referring expressions (Krauss and Weinheimer, 1966; Clark, 1996), which become progressively, contracted, systematized and abstract. This occurs for a wide range of referents, e.g. when describing spatial locations (Garrod and Doherty, 1994), music (Healey et al., 2007), conceptual structures (Schwartz, 1995; Voiklis, 2012), confidence (Fusaroli et al., 2012), temporal sequences (Mills, 2011; Verhoef et al, 2016), and also when describing how to manipulate physical objects (Shirozou, 2002). Systematization of referring expressions also occurs across modalities - in spoken interaction (Pickering and Garrod, 2004), text-based interaction (Healey and Mills, 2006) and in graphical, mediated interaction (Healey, 2001).

The development of systematicity is not simply due to the coordination problem of creating a novel referring expression: once referring expressions have been used successfully, they continue to develop (Garrod, 1999; Healey, 2004). This pattern is observed both when interlocutors are faced with the task of describing unfamiliar referents using novel referring expressions (Galantucci, 2005), as well as in situations where interlocutors already possess referring expressions and concepts that are sufficient for uniquely individuating the referents (Pickering and Garrod, 2004). Even when the names of the referring expressions are given experimentally, as in the map task (Anderson et al., 1991), interlocutors coordinate on the semantics of their referring schemas (Larsson, 2007).

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* speakers' cognitive biases - simply being exposed to another's linguistic output should yield more abstract descriptions. The interactive alignment model (Pickering and Garrod, 2004) proposes that convergence arises as a consequence of mutual priming and alignment, while the collaborative model of Clark (1996) emphasizes the role of positive feedback. One central problem with these accounts is that the basic mechanisms they propose are inherently conservative (Healey, 2004). Once a particular form is the most successfully and widely used by members of a group, there is no mechanism to explain how it might be supplanted by another. Yet interlocutors continue to develop more systematized descriptions throughout the interaction.

Further, a series of experiments (Healey and Mills, 2006; Mills and Healey, 2008) suggest that the development of abstraction can be driven by participants encountering and resolving problematic understanding. In these experiments, participants played an online version of the maze game (Pickering and Garrod, 2004) and

communicated via an experimental chat-tool which inserts artificial clarification requests into the interaction. The clarification requests appear, to participants, to originate from each other. For example in the following conversation between two participants A and B , the second turn "row?" is an artificial turn produced by the server, but appears to originate from participant B.

A: Go to the 3rd row 1st box
B: row? *(produced by the server)*
A: yeah from the top

When participants received these interventions, they produced less abstract descriptions. However, once the interventions stopped, participants subsequently used more abstract descriptions than participants who had received no interventions (Mills, 2015).

In a subsequent experiment (Healey, Mills, Eshghi, 2013) , this methodology was used to automatically detect naturally occurring clarification requests and transform them into more severe signals of miscommunication. For example in the following conversation between two participants A and B, B's clarification request "5th?" is intercepted and transformed into "what?" and sent to A.

A: go to the 5th row 2nd square
B: 5th? *(intercepted by server)*
B: what? *(transformed turn sent to B)*
A: yeah from the top

Notice that this transformation reduces the diagnostic specificity of the clarification request; A has less evidence of B's level of (mis)understanding. Since there is an expectation that a conversational partner should provide diagnostic information that is sufficient to resolve misunderstanding (Clark, 1996), this manipulation makes it appear to A that B is experiencing more difficulty than is actually the case. Participants who received these artificially amplified clarification requests also converged on more abstract descriptions than participants in a baseline condition.

Taken together, these results suggest that (1) When interlocutors encounter problematic understanding, they initially decrease the level of abstraction of their referring expressions, allowing them to identify and diagnose the nature of the misunderstanding, and (2) Once the problem has been resolved, this subsequently allows the participants to coordinate on even more abstract and systematized referring expressions.

However, these experiments have focused solely on "trouble" that is signalled in clarification requests about the content of another's turns, i.e. in "other-initiated" repair (Schegloff, 2007). It is currently unclear whether negative evidence in self-repair might also have an effect on the development of abstract referring conventions.

2 Method

To investigate in closer detail how negative evidence might contribute toward convergence, we report a variant of the maze-task. Here too, 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 left-most boundary of the edited portion of the turn and inserts a hesitation marker (e.g. "umm" or "uhhh" immediately preceding the revision), followed by the text that was deleted. 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:

- A: original turn + hesitation marker + reformulated turn
- A: original turn + hesitation marker + ''I meant'' + reformulated turn

3 Results

Interventions were performed symmetrically on both members of a dyad. No participants reported detecting 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 the transformed turns continued to use more abstract descriptions.

4 Discussion & Conclusions

The patterns observed in the maze game dialogues show that the interventions have a beneficial effect on semantic coordination. However, it is currently unclear how the constituent components of the self-repairs contributed: It could be that this effect is due entirely to the hesitation markers. Conversely, it is possible that this effect is due solely to participants reading the deleted text. If so, it is possible that the deleted text provides additional information about the other's level of (mis)understanding. It could also be that the deleted text makes the dialogue *less* coherent, forcing participants to compensate for the perturbation caused by the interventions.

Since participants encountered multiple interventions per trial, it is not possible to distinguish between the effects of the individual components. However, in aggregate we argue that 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 problems and once these problems have been identified and resolved, dyads are able to converge quicker on more stable and more abstract referring schemas.

5 References

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

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

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. Cognitive Science, 29(5), 737.

Garrod, S. C., Anderson, A., (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition, 27, 181218

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

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

Healey, P. G. (2004). Dialogue in the degenerate case? Peer commentary on Pickering & Garrod (2004). Behavioural and Brain Sciences, 27(2), 201.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure. In Simulating the evolution of language (pp. 121-147). Springer London.

Krauss, R. M., Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. JPSP, 4, 343346.

Larsson, S. (2007). Coordinating on ad hoc semantic systems in dialogue. In Proceedings of DECALOG

Mills, G. J. (2011). The emergence of procedural conventions in dialogue. Proceedings of the Cognitive Science Society (pp. 210-211).

Pickering, M. J., Garrod, S. (2004).Towards a mechanistic psychology of dialogue. Behavioural and Brain Sciences, 27(2), 169190.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: 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.

Shirouzu, H.,, Miyake, N., Masukawa, H. (2002). Cognitively active externalization for situated reflection. Cognitive science, 26, 469501

Verhoef, T., Walker, E., Marghetis, T., (2016) Cognitive biases and social coordination in the emergence of temporal language. Proceedings of Cog Sci