Quantitative Models of Human-Human Conversational Grounding Processes

Clayton D. Rothwell (crothwell@infoscitex.com)
Infoscitex Corp., 4027 Colonel Glenn Hwy
Wright State U. Department of Psychology, 3640 Colonel Glenn Hwy
Dayton, OH 45435 USA

Valerie L. Shalin (valerie.shalin@wright.edu)
Wright State U. Department of Psychology and Kno.e.sis, 3640 Colonel Glenn Hwy
Dayton, OH 45435 USA

Griffin D. Romigh (griffin.romigh@us.af.mil)
Air Force Research Laboratory, 2610 Seventh St
Wright-Patterson AFB, OH 45433

Abstract

Natural language dialogue between multiple participants requires conversational grounding, a process whereby interlocutors achieve a shared understanding. However, the mechanisms involved in the grounding process are under dispute. Two prominent models of dialogue between multiple participants are: *interactive alignment*, a simpler model that relies on automatic priming processes within individuals, and *interpersonal synergy*, a more complicated model emphasizing coordinated interaction across participants. Using recurrence analysis methods, Fusaroli and Tylén (2016) simultaneously evaluated both models and showed that alignment is an insufficient explanation for grounding or for the teams' task performance. However, their task and resulting dialogues lack the typical complexity of conversations or teamwork. Furthermore, the interpersonal synergy model was not clearly differentiated from other coordination-focused models of grounding with explicit foundations in strategy and intentionality (i.e., audience design, joint activity, perspective taking). Here we test recurrence-based models in a collaborative task that stressed the grounding process. Results support a coordination model of dialogue over the alignment model as a predictor of performance. Content-based mediation analyses showed that the coordination recurrence model includes critical aspects of strategic design and is not purely interpersonal synergy.

Keywords: Communication; Dialogue; Conversational Grounding; Multi-person Cognitive Models

Introduction

The grounding process is a key focus of human dialogue research (Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991; Branigan, Pickering, & Cleland, 2000; Pickering & Garrod, 2004). Conversational grounding is the process whereby interlocutors determine that they have understood one another, and results in additions to shared knowledge and understanding. Grounding underlies successful collaboration by developing a shared context, supporting immediate feedback of actions, and allowing for incremental progress in conveying intent (Brennan, 1998). Grounding processes influence communication effectiveness and resulting performance metrics such as laboratory task completion time (Clark & Wilkes-Gibbs, 1986; Clark & Krych, 2004; Reitter & Moore, 2014), due in part to the requirements for shared understanding in collaborative tasks.

The proposed models of dialogue use multiple conversational participants as the unit of analysis, but suggest different mechanisms for grounding processes (Horton & Gerrig, 2005; Louwerse, Dale, Bard, & Jeuniax, 2012; Schober & Brennan, 2003). One prominent model is *interactive alignment* (Pickering & Garrod, 2004). Alignment (also entrainment, convergence or imitation) refers to the increasing similarity of the interlocutors through adoption of each other's phonetic, prosodic, lexical, or syntactic content (Branigan et al., 2000). Alignment credits this to priming, an automatic, covert mechanism by which recent experiences influence the likelihood of future contributions. Alignment at lower linguistic levels presumably propagates to the semantic level and the situation model of the interlocutors, which forms the basis of a shared understanding of each other and of the world. Thus alignment provides an appealing, conceptually straightforward explanation of grounding phenomena.

Some researchers question the sufficiency of alignment to explain common ground and grounding of new material. The prominent alternative models of human grounding processes emphasize coordination and complementarity such as adjacency pairs in the interaction, rather than similarity. *Coordination* models separate into two categories: *strategic design* and *interpersonal synergy*. In *strategic design*, speakers appear to design utterances in light of their audience's knowledge. The knowledge may concern the audience's culture, group membership, spatial perspective, or previous conversational interactions (Clark & Marshall, 1981; Schober, 1993). Strategic design is marked by intentionality—goal-directed conversational behavior that seeks and displays evidence of understanding. These goals invoke an additional layer of exchange concerning the collaborative management of the conversation, called Track 2 dialogue (Clark, 1996). Track 2 dialogue includes: acknowledgements of understanding, displays of non-understanding, and requests for clarification. *Interpersonal synergy* is a recent and relatively less examined theory. The coordination from interpersonal synergy either does not require intentionality (Fusaroli, Raczaszek-Leonardi, & Tylén, 2014), or redefines it (see also Gallagher & Miyahara, 2012). Interlocutor coordination emerges from a complex dynamical system achieving stability in a specific context, and becomes cemented in interaction routines. The introduction of new interlocutors into established interaction routines disrupts communication (Fusaroli et al., 2014).

Quantification of Recurrence

Separate bodies of research have investigated interactive alignment and coordination models, but to the authors' knowledge only one study has attempted to examine the two theories with competing models for the same performance data (Fusaroli & Tylén, 2016). Recent advancements in the analysis of dyadic dialogue utilize the non-linear analysis methods of recurrence quantification analysis (RQA) and cross recurrence quantification analysis (CRQA). RQA and CRQA originated from the study of dynamic systems and were developed to examine recurrence in chaotic systems, i.e., repetition of states in time series data. RQA seeks recurrence within one time series (analogous to autocorrelation) and CRQA seeks recurrence between two time series (analogous to cross-correlation). These methods reveal and quantify order and organization that is not readily apparent. Originally built for continuous data, these methods have been adapted for categorical data and used in the analysis of lexical content (e.g., Orsucci et al., 2013) and syntax (e.g., Dale & Spivey, 2006).

Grounding Process Models

Fusaroli and Tylén (2016) created models for alignment and coordination using recurrence analyses, and discriminated between these models by their relationship to task performance. They argued that the coordination recurrence model specifically represented interpersonal synergy, though strategic design is an unexamined alternative. Their approach is illustrated in Figure 1. The same time series contents appear in each panel but different outlined patterns reflect the different recurrence sensitivities. The alignment model detects patterns transferred from one speaker to the other, such as 'XYY' from speaker A to speaker B (though not illustrated, patterns that go from B to A will also be detected). The coordination model detects speaker-independent patterns, including patterns across speakers such as adjacency pairs. For instance, the pattern 'YXZXY' occurs between A and B and later B and A. In the self-consistency baseline model, recurrence of patterns within A and within B were tested separately selecting for analysis whoever had the higher recurrence rate.

The dialogue in Fusaroli and Tylén (2016) resulted from two participants performing a visual detection task. Each participant made an independent judgment of whether the target signal appeared in the first interval or the second interval of the stimulus. Dialogue only occurred when their judgment disagreed—they discussed the stimulus and came to a collaborative judgment. Collaborative benefit was computed as the ratio between joint performance and the highest individual's performance, where ratio values greater than 1 indicated a benefit from the joint decision. Recurrence values for lexical choice, pauses, and prosody were calculated according to each theory and then used as predictors in separate regression models with collaborative benefit as the outcome, thereby relating each grounding model to task performance.

Both the alignment and coordination models were related to task performance, but coordination was a better predictor of performance for the lexical level and the speech/pause level. The two models were similar for the prosodic level.

This quantitative approach provides a promising beginning to the direct comparison between grounding models. However, the task and dialogue content was very limited. The task used simple visual psychophysics stimuli and required a simple choice between two intervals. The vocabulary and conceptualizations that appeared in the dialogues, though not reported, were most likely very limited. The importance of these results for more complex dialogues in a more complicated task setting was not established. In addition, the analysis failed to differentiate between interpersonal synergy and strategic design. Although similar in their emphasis on coordination, these models maintain important distinctions regarding the characterization of cognitive mechanisms. Intentionality (and goal-directed behavior) is one way to differentiate between the two models but their correspondence to the coordination recurrence model is not intuitive. Additional analyses must distinguish between synergy on the one hand and design and intentionality on the other.

Current Study

The current study examined a team task that stressed the grounding process and applied the RQA and CRQA models for coordination and alignment on two lexical levels: the morpheme level used in Fusaroli and Tylén (2016) and the word level. As discussed below, word-level analysis facilitated an additional mediation analysis of recurrence model results. The task and resulting team dialogue resulted in rich, long dialogue with numerous and diverse content to stress grounding processes. Consistent with Fusaroli and Tylén, we hypothesized that the recurrence metrics calculated based on the coordination model would have a stronger relationship to performance than the alignment model. In addition, we sought to investigate what the coordination recurrence model is measuring, and its relationship to the strategic design model of grounding. We created a lexicon of Track 2 dialogue (described below) and tested if the coordination model statistically mediates the relationship between Track 2 dialogue and performance. Mediation can identify a process that underlies an observed relationship (Baron & Kenny, 1986). We used mediation to see if the variance in performance explained by Track 2 dialogue is reduced by the coordination model. Such mediation demonstrates that the recurrence model for coordination captures aspects of strategic design in addition to, or possibly instead of, interpersonal synergy.

Methods

Uncertainty Elicitation Task Corpus

We used materials from the Uncertainty Elicitation Task corpus (Romigh, Rothwell, Greenwell, & Newman, 2016).

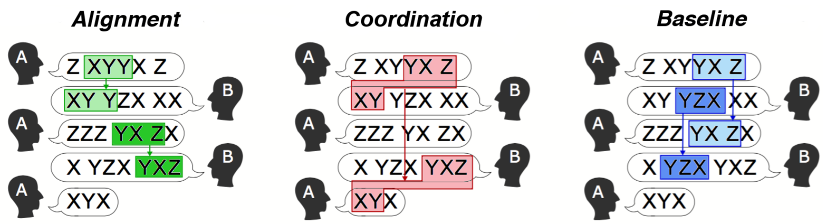

Figure 1: Illustration of the recurrence tests for alignment, coordination, and baseline (adapted from Fusaroli & Tylén, 2016). Alignment models were sensitive to patterns transferred between speakers. Coordination models were sensitive to patterns independent of speaker, which included patterns across speakers as illustrated here. Baseline models were sensitive to patterns within one speaker (i.e., self-consistency). (Figure used with permission from John Wiley and Sons).

Like Fusaroli and Tylén (2016), this was a symmetric dialogue task—no one speaker had the answer, so the conversational dynamics were flexible and negotiable. Partners had many unlabeled pictures of various real world scenes from both an overhead perspective and street-level perspectives, that they had to match with each other. This led to conceptually complex and diverse dialogues. Partners discussed: house features (e.g., siding, roof, windows, garage, porch, columns, 1 or 2-story), lot features (e.g., trees, yard, fence, driveway, garden, sidewalk, corner lot, playground, pool), street/neighborhood features (e.g., presence of stop-sign, power lines, presence of alleyways, nearby parks), and car features (e.g., number of cars, type of vehicle: truck/van/sedan, color).

On each trial, two people sat in separate rooms and worked together to locate street-level pictures of different houses on an overhead map (Figure 2). Street-level images and satellite images were obtained from Google Maps with labels (e.g., street names) removed. The overhead map was the same for both participants and had 12 numbered buildings (1-12). The participants each had street-level pictures of 6 of those buildings on the right hand side of their screens. The participants were given street-level images from different points of view and they had to determine that they were discussing the same building. Their task was to relate the street-level views to the overhead map by labeling the street-level with a number 1-12, and the trial ended when all street-level images had correct number labels.[1] As accuracy was held constant, completion time was the performance metric (shorter times indicated better performance). Performance in the task was expected to be related to conversational grounding because participants needed to communicate effectively—make definite references to unlabeled street-level views of houses, share the information from their unique street-level views, and discuss the similarity between street-level and overhead imagery. Five teams of 2 people and each team completed 8 trials for a total of 40 trials.[2]

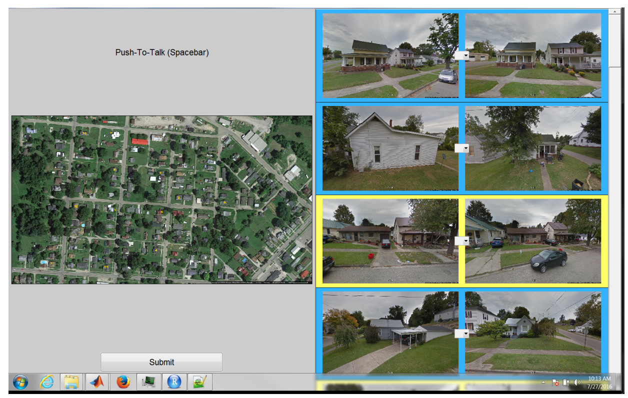

Figure 2: Screen shot from the Uncertainty Elicitation Task. Building numbers appear on the overhead map and participants labeled street-level images using the drop-down boxes centered on each row of images.

Recurrence Analyses

Our analysis examines recurrence at two lexical levels (the word level and the morpheme level) in search of the model that best predicts task-specific performance.

Prior to calculating the recurrence plot, RQA and CRQA require a number of parameters. We used values keeping with Fusaroli and Tylén (2016) and other categorical analyses of transcript data (Orsucci et al., 2013). The radius value was set to 0, meaning only an exact match would be counted as a recurrence, which is appropriate for nominal data. For example, for the word-level analysis each word was given an arbitrary unique numerical identifier. The threshold for a line (i.e., recurrence patterns that are parallel to the positive diagonal) was set at 2. Time delay was set to 1. The word-level analysis used single words as the unit of analysis, which was specified by an embed value of 1. The morpheme level[3] used a 3-letter unit of analysis (i.e., a letter trigram), which was specified by an embed value of 3.

[1] Other experimental manipulations in the original corpus were not the focus of our analysis.

[2] To address the within-subjects nature of the data, we removed variance due to teams in a secondary analysis (summarized below) and found a similar pattern of results. We did not test for order effects or learning, but if learning has occurred it would increase the performance variability in the dataset.

[3] Fusaroli & Tylén's *lexical choices*

The three models in Figure 1 were tested. The *alignment* model was represented by CRQA of a time series of Speaker A with a time series of Speaker B. To preserve the time sequence and phase information of the entire dialogue, added codes in each time series replaced the other speaker's contributions. The *coordination* model was represented by RQA of the time series for the entire block (Speakers A & B). A baseline *self-consistency* model was represented by performing RQA of each speaker's time series with his/herself and using the recurrence plot with the highest recurrence rate.

The three separate recurrence models output separate recurrence metrics for different regression models, in order to assess the relationship of each recurrence model to task performance. This analysis process differs from a typical regression procedure where predictors are added or removed from a single regression model. Here, three different regression models with the same four predictors were based on different recurrence calculations. Individual recurrence metrics of recurrence rate, determinism, average line length, and line entropy were calculated for each of the three models (i.e., alignment, coordination, baseline), for each trial in the Uncertainty Elicitation Task corpus. The recurrence metrics quantify how much recurrence occurs (*Recurrence Rate, RR*), the proportion of recurrence that appears in longer sequences (*Determinism, DET*), the average length of recurrence sequences (*Line Length, L*), and the variety in recurrence lengths (*Line Entropy, ENT*). Recurrence metrics then functioned as predictors for a linear regression model of the performance scores (i.e., completion times) for each trial. Linear models were evaluated using $AdjR^2$ values. These models follow the analysis procedures from Fusaroli and Tylén (2016) assuming data from a between-subjects design. Subsequent tests address the repeated measures (i.e., within team) nature of our data. Analyses were performed in R, using the *crqa* package (Coco & Dale, 2014).

We also created a lexicon (i.e., word list) for Track 2 dialogue using the Linguistic Inquiry and Word Count 2015 (LIWC) text analysis program (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Our analysis relied on two separate lexicons that may capture Track 2 issues of dialogue management: *Assent* (e.g., agree, OK, yes) and *Certainty* (e.g., indeed, always, never). We reasoned that *Assent* may capture an addressee's acceptance of a speaker's installment and *Certainty* may capture confusion regarding an installment. We tested how well the LIWC categories accounted for performance by using the LIWC counts (frequencies of words in the *Assent* and *Certainty* lists) for each trial as predictors of performance. We then tested if the parameters from the coordination recurrence model statistically mediated the relationship between the LIWC categories and performance, following the 'Causal Steps' procedure (Baron & Kenny, 1986). This involved three "Steps" where the LIWC categories were treated as independent variables (IVs) and the recurrence parameters were treated as mediators (Ms): 1) the IVs and performance, 2) the IVs and the Ms, and 3) the (IVs + Ms) and performance. Multiple linear regression was used for Steps 1 and 3 while MANOVA was used for Step 2 in order to test for a relationship between multiple LIWC categories and multiple recurrence-parameter mediators.

Results

First we show that the observed recurrence was not due to chance. Next we show that the coordination model has stronger relationships to task performance than the alignment model for both the word-level and the morpheme-level analyses. Moreover, the coordination model accounts for variance in performance after controlling for team differences whereas the alignment model does not. Mediation analysis shows that the coordination model reflects aspects of Track 2 dialogue.

Chance Analysis

The structure of recurrence represented by these metrics was not due to chance. We compared the outputs of the recurrence analyses of the data to outputs using a shuffled time series (i.e., randomly ordering the words in the time series).

Paired *t*-tests indicated that recurrence structure is significantly different from shuffled controls for all models for all values (all = $p < .0001$), except for word-level recurrence rate. For the word-level test, the recurrence rates are exactly the same because shuffling does not add or remove words.

Word-Level Analyses

The linear regression models for word-level analyses are shown on the left side of Table 1. The coordination model accounted for more of the variance in completion times ($AdjR^2$ = 0.66) than the alignment model and the baseline model ($AdjR^2$ = 0.14 & 0.51, respectively). The baseline model accounted for more variance than the alignment model.

Morpheme-Level Analyses

The linear regression models for morpheme-level analyses are shown on the right side of Table 1. The pattern of results was the same as the word level. The coordination model accounted for more of the variance in performance ($AdjR^2$ = 0.76) than the alignment model or the baseline model ($AdjR^2$ = 0.32 & 0.64, respectively). The baseline model accounted for more variance than the alignment model as well.

Controlling for Team Differences

Space precludes a complete presentation, but controlling for team differences was necessary for the within-subjects design and Team ID was a significant predictor of performance ($F(4, 35) = 6.04$, $p < .001$, $AdjR^2 = 0.34$). Using statistical controls that removed the variance between teams, we tested if each recurrence model could explain the residual variance. The alignment model did not ($F(4, 31) = 0.53$, $p = .72$, $\Delta R^2 = 0.04$) whereas the coordination model did ($F(4, 31) = 8.21$, $p < .001$, $\Delta R^2 = 0.30$).

Table 1: Word-level analyses (left) and Morpheme-level analyses (right)—linear regression models for alignment, perspective taking, and baseline. Predictors were recurrence rate (RR), determinism (DET), average line length (L), and line entropy (ENTR). (*$p < .05$, **$p < .01$, ***$p < .001$)

	Word-Level				Morpheme-Level			
Theory	**Contents**	***p*-value**	**Contents**	***p*-value**	**Contents**	***p*-value**	**Contents**	***p*-value**
Alignment*	$\boldsymbol{AdjR^2 = 0.14}$	**0.051**			$\boldsymbol{AdjR^2 = 0.32}$	$\boldsymbol{< .01}$		
	RR_A	0.53	L_A*	$< .05$	RR_A	0.21	L_A*	$< .05$
	DET_A	0.13	$ENTR_A$**	$< .01$	DET_A***	$< .001$	$ENTR_A$*	$< .05$
Coordination***	$\boldsymbol{AdjR^2 = 0.66}$	$\boldsymbol{< .001}$			$\boldsymbol{AdjR^2 = 0.76}$	$\boldsymbol{< .001}$		
	RR_S**	$< .01$	L_S***	$< .001$	RR_S	0.26	L_S***	$< .001$
	DET_S	0.48	$ENTR_S$	0.07	DET_S***	$< .001$	$ENTR_S$***	$< .001$
Baseline***	$\boldsymbol{AdjR^2 = 0.51}$	$\boldsymbol{< .001}$			$\boldsymbol{AdjR^2 = 0.64}$	$\boldsymbol{< .001}$		
	RR_B	0.17	L_B	0.15	RR_B***	$< .001$	L_B***	$< .001$
	DET_B**	$< .01$	$ENTR_B$	0.52	DET_B*	$< .05$	$ENTR_B$**	$< .01$

Mediation Analysis

Linear regression results for the LIWC categories *Assent* and *Certainty* appear at the top of Table 2. These categories significantly predicted task completion times ($AdjR^2$ = 0.43). The coefficients for *Assent* and *Certainty* were both negative. More instances of these words resulted in faster completion times (i.e., better performance). Additional mediation tests used the word-level coordination model's recurrence parameters. The MANOVA in Step 2 showed that the LIWC lists were related to these recurrence parameters ($F(4, 34) = 6.62$, $p < 0.001$, and $F(4, 34) = 9.70$, $p < 0.001$, respectively). Step 3 showed that the recurrence parameters mediated the relationship between task completion times and LIWC categories (Table 2 bottom portion) by eliminating their significance.

Table 2: Mediation analysis for LIWC—See text for details. (*$p < .05$, **$p < .01$, ***$p < .001$)

Step 1— LIWC's relation to performance			
Assent***	$< .001$	Certainty**	$< .01$
Step 2— LIWC's relation to Coordination			
Assent***	$< .001$	Certainty***	$< .001$
Step 3—LIWC's & Coordination's relation to performance			
Assent	0.30	Certainty	0.89
RR_S*	$< .05$	L_S**	$< .01$
DET_S	0.32	$ENTR_S$	0.32

Discussion

Findings clearly supported the coordination model over the alignment model for both levels of analysis. At the word level, coordination accounted for 52% *more* of the variance in task completion times than alignment, and 44% *more* at the morpheme level. Although the baseline model performed better than Fusaroli and Tylén (2016), the pattern of findings for alignment and coordination was similar. Moreover, the relationships between coordination and performance found here were larger than those shown by Fusaroli and Tylén, despite the longer, more complex dialogues. While the coordination model accounted for performance above team differences, the alignment model did not.

Recent research agrees with these findings that communication processes are more complicated than priming-based alignment. Rather than repeating content, interlocutors' contributions provide new content that compliments past contributions (Tenbrink, Andonova, & Coventry, 2008). Many studies of alignment do not include performance outcomes (e.g., Branigan et al., 2000) and therefore may not identify these insufficiencies. Alignment may still occur over longer time scales, which has been shown to predict task performance (Reitter & Moore, 2014). The alignment recurrence model used here did not distinguish between short-term and long-term alignment, so it is possible that long-term alignment is responsible for the relationship between alignment and performance.

Beyond support for a general coordination model, the coordination recurrence model appears to contain aspects of strategic design. Indeed, Track 2 dialogue alone accounted for more variance in performance than alignment did at both the word level and morpheme level ($AdjR^2$ = 0.43 vs. 0.14 & 0.34, respectively). Recent research supports the importance of design—utterances often reflect different perspectives and interlocutors appear to keep track of multiple perspectives at the same time (Brennan, Schuhmann, & Batres, 2013).

Conclusion

In this paper, we quantitatively modeled conversational grounding processes between two interlocutors. We tested two models for this process, *alignment* and *coordination*, in a complex collaborative grounding task. The results clearly discount an alignment model as a sufficient model of the conversational grounding process. Results also indicated that the coordination recurrence model is closely related to Track 2 dialogue and therefore strategic design models of the conversational grounding process must be considered. Our future research will examine whether strategic design accounts for these findings in addition to or to the exclusion of interpersonal synergy.

Acknowledgments

This research was partially supported by the Air Force Research Laboratory's Human Interface Research Technologies Contract #: FA8650-14-D-6501. We thank W. Sid Horton for valuable discussions and Kno.e.sis for use of the Linguistic Inquiry and Word Count software.

References

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–25. doi: 10.1016/S0010-0277(99)00081-5

Brennan, S. E. (1998). The Grounding Problem in Conversations With and Through Computers. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 201–225). Hillsdale, NJ: Lawrence Erlbaum.

Brennan, S. E., Schuhmann, K. S., & Batres, K. M. (2013). Collaboratively Setting Perspectives and Referring to Locations Across Multiple Contexts. In *Proceedings of the PRE-CogSci 2013 Workshop on the Production of Referring Expressions* (Vol. 1, pp. 1–6). Berlin, Germany.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in Communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, D.C.: American Psychological Association. doi: 10.1037/10096-006

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. doi: 10.1016/j.jml.2003.08.004

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39. doi: 10.1016/0010-0277(86)90010-7

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Psychology*, *5*(June), 1–14. doi: 10.3389/fpsyg.2014.00510

Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, *56*(3), 391–430. doi: 10.1111/j.1467-9922.2006.00372.x

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*(1), 147–157. doi: 10.1016/j.newideapsych.2013.03.005

Fusaroli, R., & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171. doi: 10.1111/cogs.12251

Gallagher, S., & Miyahara, K. (2012). Neo-pragmatism and Enactive Intentionality. In J. Schulkin (Ed.), *Action, perception and the brain* (pp. 117–146). Basingtoke, UK: Palegrave-Macmillan.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142. doi: 10.1016/j.cognition.2004.07.001

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniax, P. (2012). Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, *36*(8), 1404–1426. doi: 10.1111/j.1551-6709.2012.01269.x

Orsucci, F., Petrosino, R., Paoloni, G., Canestri, L., Conte, E., Reda, M., & Fulcheri, M. (2013). Prosody and synchronization in cognitive neuroscience. *EPJ Nonlinear Biomedical Physics*, *1*(1), 1:6. doi: 10.1140/epjnbp13

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015* (Tech. Rep.). Austin, TX: University of Texas at Austin. doi: 10.1068/d010163

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190–226. doi: 10.1017/S0140525X04000056

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. doi: 10.1016/j.jml.2014.05.008

Romigh, G., Rothwell, C., Greenwell, B., & Newman, M. (2016). Modeling uncertainty in spontaneous speech: Lexical and acoustic features. *The Journal of the Acoustical Society of America*, *140*(4), 3401-3401. doi: 10.1121/1.4970912

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1–24. doi: 10.1016/0010-0277(93)90060-9

Schober, M. F., & Brennan, S. E. (2003). Processes of Interactive Spoken Discourse: The Role of the Partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Routledge.

Tenbrink, T., Andonova, E., & Coventry, K. (2008). Negotiating spatial relationships in dialogue: The role of the addressee. In *Proceedings of LONDIAL - The 12th SEMDIAL Workshop* (pp. 193–200).