Predictions of Miscommunication In Verbal Communication During Collaborative Joint Action

Alexandra Paxton[1,2†], Jennifer M. Roche[3†], Alyssa Ibarra[4], & Michael K. Tanenhaus[5]

[†]Authors Paxton and Roche equally contributed to this work.

[1]Department of Psychological Sciences, University of Connecticut

[2]Center for the Ecological Study of Perception and Action, University of Connecticut

[3]Department of Speech Pathology & Audiology, School of Health Sciences, Kent State University

[4] Department of Brain & Cognitive Sciences, University of Rochester

[5]Department of Brain & Cognitive Sciences, University of Rochester

Conflicts of Interest. The authors have no conflicts of interest to report.

Funding Statement. Preparation of this manuscript was supported by a National Institute of Health grant (R01 HD027206) to M. K. Tanenhaus.

Corresponding Author:

Alexandra Paxton

Department of Psychological Sciences

University of Connecticut

Storrs, CT, 06269

Email: alexandra.paxton@uconn.edu

Abstract

Purpose: The purpose of the current study was to examine the lexical and pragmatic factors that may contribute to turn-by-turn failures in communication (i.e., miscommunication), that arise regularly in interactive communication.

Method: Using a corpus from a collaborative dyadic building task, we investigated what differentiated successful from unsuccessful communication and potential factors associated with the choice to provide greater lexical information to a conversation partner.

Results: We found that more successful dyads' language tended to be associated with greater lexical density, lower ambiguity, and fewer questions. We also found participants were more lexically dense when accepting and integrating a partner's information (i.e., *grounding*) but were less lexically dense when responding to a question. Finally, an exploratory analysis suggested that dyads tended to spend more lexical effort when responding to an inquiry and used assent language accurately—that is, only when communication was successful.

Conclusion: Together, the results suggest that miscommunication both emerges and benefits from ambiguous and lexically dense utterances.

Keywords: miscommunication, dialogue, ambiguity, conversation, grounding

Predictors of Miscommunication in Verbal Communication During Collaborative Joint Action

Miscommunication—that is, the failure to communicate an intended message to another person—is often seen as an unfortunate byproduct of everyday communication. It has been blamed for a host of negative short- and long-term effects on communication, from creating momentary discomfort to damaging interpersonal relationships (e.g., Guerrero, Andersen, & Afifi, 2001; Keysar, 2007; McTear, 1991; 2008). Given these harmful effects, psycholinguistic research on miscommunication has tended to focus on understanding how communication breakdowns are repaired (Bazzanella & Damiano, 1999; Levelt, 1983).

However, there is currently little understanding of the processes of miscommunication itself. Although many domains that are visibly affected by miscommunication explored the negative effects of miscommunication, understanding how miscommunication works—and even how we might be able to use it to our advantage—may help us mitigate communication failure. Research in healthcare-related fields has shown alarming effects of miscommunication on patient health. Unfortunate and even fatal recovery outcomes have been linked to miscommunications about care between caregivers and surgical patients (Halverson et al., 2011; Lingard et al., 2004). An estimated 15.8% of medication errors stem from miscommunication about appropriate use (Phillips et al., 2001), and approximately 32% of unplanned pregnancies are related to miscommunications about effective contraception use (Isaacs & Creinin, 2003). Perhaps most alarmingly, 67% of trauma patient deaths result directly from miscommunication between members of the trauma team (Raley et al., 2016); in 2000 alone, between 44,000 and 98,000 people died in hospitals because of medical miscommunication (Sutcliffe, Lewton, & Rosenthal, 2004). These efforts underscore the potential for direct application of basic research into the processes of miscommunication to improve lives.

Most consequences of miscommunication are not this dire, but these examples demonstrate the importance of studying miscommunication. A thorough understanding of miscommunication cannot simply propose methods to prevent it but must also improve our understanding of how we function despite it. Before we can promote ways to prevent the most severe negative consequences of miscommunication, we must build a foundation for understanding how miscommunications occur in language during interaction. In the current study, we contribute to the basic study of miscommunication by examining its pragmatic and lexical contributors within a collaborative task.

Miscommunication as an Opportunity for Success

Previous work on learning has suggested that learning may be more likely to happen when the cognitive system is perturbed, thanks to the recruitment of additional attentional resources (D'Mello & Graesser, 2011; Graesser & Olde, 2003). This raises the possibility that miscommunication may sometimes provide a stepping-stone for improved communication: Miscommunication can capture attention when it perturbs the cognitive system by triggering the learner or listener to recruit attentional resources to the situation.

Successful communication necessarily requires interlocutors to coordinate and regularly update their mutual knowledge, experiences, beliefs, and assumptions (e.g., Clark & Carlson, 1982; Clark & Marshall, 1981). One way that interlocutors can do this is by establishing *conceptual* or *lexical pacts*, negotiating meanings of shared items or experiences with one another (Brennan & Clark, 1996). These pacts may not always be explicit (cf. Fusaroli et al., 2012; Mills, 2014), but these shared ideas and referential expressions quickly coordinate joint action. However, the *grounding* process—that is, the process of establishing these pacts—is often riddled with unsuccessful attempts that slowly pave the way to a common goal. Some researchers have provided insights into how interlocutors might resolve communication problems (e.g., through ambiguity resolution, asking clarification questions, and repair; Clark & Brennan, 1991; Garrod & Pickering, 2004; Haywood, Pickering, & Branigan, 2005; Levelt & Cutler, 1983). Interlocutors must therefore approach conversations with relative flexibility to adapt to moment-to-moment changes in conversational demands in order to successfully negotiate shared activities (Ibarra & Tanenhaus, 2016).

At the same time, interlocutors do not want to provide more information than necessary (e.g., Grice, 1975). Increased information can tax the listener's cognitive resources and can result in inappropriate inferences. Producing the additional information will also be costly for the talker. By investing effort when important new information is introduced during the interaction, interlocutors can work together to establish efficient pacts by more equitably distributing effort (even implicitly; Brennan & Clark, 1996; Zipf, 1949).

During extended collaborative dialogue, what appears to be under-specification—that is, where the talker appears to be giving less information in a given utterance than is often needed to uniquely refer—is quite common: Because talkers' referential domains become closely aligned through their interaction, seemingly under-informative referential expressions actually provide necessary and sufficient information in the context of their shared goals and task constraints (Brown-Schmidt & Tanenhaus, 2008). However, problems may arise when a talker inaccurately estimates the listener's needs or the pair's conceptual pacts, goals, and task constraints.

Therefore, interlocutors must delicately balance when they must provide additional information and when they can get away with saying as little as possible. If a talker is too "cheap" in their message, the omission of critical details could lead the interaction to suffer. On the other hand, if a talker's message is too "expensive," heavy cognitive demands may cause the interaction to suffer, including interlocutors making unnecessary and even inappropriate inferences. In fact, ambiguity may even be a feature (not a flaw) of communication to maximize efficiency so long as the context is sufficiently rich (Piantadosi, Tily, & Gibson, 2012).

When reducing effort by providing less information, ambiguous language is likely to increase. However, listeners expect reduced information under some circumstances; for example, a "repeated name penalty" occurs when a talker repeats a name when a pronoun is expected (Gordon, Grosz, & Gilliom, 1993). In fact, using a fully specified referent—regardless of the state of discourse—increases processing difficulty relative to language with potentially ambiguous referents (Campana, Tanenhaus, Allen, & Remington, 2011).

Because spoken language unfolds over time, listeners routinely encounter temporary ambiguity at the segmental, lexical, and syntactic levels. When a talker uses ambiguous language, the listener may be able to situate it within the current context and easily settle on the talker's meaning. To reduce some of the burden placed on a single individual's cognitive system, interlocutors may communicate more easily by offloading some of the processing effort to one another and to the broader interaction context (e.g., Zipf, 1949).

However, listeners may not always understand the intended message from an ambiguous reference, leading to moments of uncertainty and misinterpretation. At this point, communication does not necessarily fail entirely. Instead, various processes within the dyadic system allow the listener to confirm the talker's intent and solicit more information when the message is unclear. For example, back-channeling—or brief responses from the listener during a speaker's turn—can increase conversational flow between interlocutors and indicate that the listener understands the speaker (Bavelas & Gerwing, 2011; Lambertz, 2011; Yngve, 1970).

We cannot always know when our referential domains are completely aligned and when they have become mismatched. An efficient strategy, then, may be to provide utterances that are as minimally "content-full" (or lexically dense) as needed by the current context. However, with such a strategy, unless interlocutors' referential domains are *perfectly* aligned throughout an entire interaction, miscommunication will likely follow from missing or impoverished information, at least occasionally. We can view this strategy as arising from interlocutors' attempts to balance talker effort with listener understanding in an uncertain environment.

Given this view, efficient task-oriented dialogue should be marked by intermittent instances of miscommunication. These would likely occur when language is just a bit too ambiguous or missing just a bit too much information. Under this view, miscommunication should be both common and a natural consequence of minimizing communicative effort, with interlocutors providing additional information only when prompted by miscommunication.

The Present Study

Previous psycholinguistic research has demonstrated how pragmatic and linguistic behaviors impact language processing. We aim to contribute to this literature by quantifying the roles that a targeted subset of pragmatic and lexical behaviors plays in miscommunication. More closely evaluating the behaviors associated with miscommunication may shed light on the processes behind miscommunication. At present, miscommunication is poorly understood, but it is likely tied to basic cognitive processes and patterned aspects of the communicative context.

We created an interactive dyadic task with a clear turn structure with an objective measure of communicative success. Crucially, partners had to work together toward a shared goal without a shared visual environment, allowing us to specifically target the contributions of language to performance and miscommunication. The task allowed us to hold overall success constant: Because all dyads eventually completed the joint task successfully, we could separate the dynamics of local success (i.e., the turn-by-turn successes or miscommunications) from global success (i.e., achieving the stated goal of the interaction). Rather than examining overall success or confounding overall and local success, we were able to look at how each dyad's moment-to-moment success or failure were related to their language patterns. By operationalizing local miscommunication and restricting communication to explicit linguistic patterns, we were able to isolate specific contributions to communicative success or failure.

Through experimental paradigms like the map task (e.g., Anderson et al., 1991) or the tangram task (e.g., Clark & Wilkes-Gibbs, 1986), researchers have built decades of findings on the ways in which interacting individuals emerge from miscommunication during joint action through the constellation of studies on *repair*. We seek to complement these findings by explicitly focusing on the characteristics of miscommunication itself. By directly comparing successful and unsuccessful communication, we can better understand the processes of communication more broadly. To do this, we consider the roles of linguistic and pragmatic behaviors in "local" (or turn-by-turn) miscommunication.

How Pragmatic and Lexical Behaviors Affect Local Miscommunication (Model 1). Miscommunication may emerge as a result of the (mis-)interpretation of pragmatic behaviors and lexical items within the specific conversational context. We target five pragmatic and lexical behaviors that could contribute to turn-by-turn failures in communication: the use of task-specific ambiguous language, the use of statements of assent or negation, responding to a question, and the amount of content being conveyed between interlocutors (operationalized here as lexical density; see Measures section). These behaviors—while individually interesting and vital to successful communication—may together influence the dynamics of turn-level success.

By its nature, ambiguous language omits concrete or explicit content; therefore, if that ambiguous utterance is not sufficiently grounded, miscommunication is likely to follow. Although ambiguity can emerge naturally from a variety of sources (e.g., increased cognitive load, assumed grounding, failures in perspective-taking), we are here able to isolate ambiguous language in a task-relevant domain: spatial terms. Since partners lack a shared visual environment in our task, any spatial referent will be somewhat ambiguous, allowing us to examine how these behaviors influence miscommunication.

Questions are an essential pragmatic behavior, allowing interlocutors to request clarification or to check if their partner requires clarification. Whether an interlocutor is responding to a question could provide useful information about the pragmatic state of the conversation, even when ignoring the semantics. Under the current assumption that interlocutors may be prompted to include more detail only when asked a question by their partner, we choose here to focus on *responses* to questions (rather than to questions themselves).

In spite of the "yes" bias (i.e., the increased likelihood of individuals to answer a question with an affirmation rather than a negation; e.g., McKinstry, Dale, & Spivey, 2008) and the tendency to back-channel using affirmations (rather than negations or other types of words; e.g., Schegloff, 1982), individuals should be more likely to use assent words to establish grounding or signal understanding within this context. Similarly, interlocutors should be more likely to use negation when communication falters (e.g., when aware of their own lack of understanding).

Finally, interlocutors should only provide one another with the information necessary within the conversational context (Grice, 1975). However, interlocutors may have difficulty providing the appropriate amount of information when deprived of vital shared information within the conversation context—including a shared visual environment, as in the current study. Given the difficulties associated with these pressures, we hypothesize that miscommunication will be associated with content-impoverished (i.e., lexically shallow) utterances as compared with content-rich (i.e., lexically dense) utterances.

Taken together, we hypothesize that increased use of ambiguous language, negation, and lexically shallow utterances will be associated with miscommunication in a given turn—all of which may stem from the difficulty in accurately providing the amount and type of content needed to promote success. However, we hypothesize that assent, responding to a question, and more lexically dense utterances will predict successful communication in a given turn.

How Joint State and Pragmatics Shape Communication Richness (Model 2). We are also interested in identifying the circumstances in which interacting individuals provide their partners with additional information. Certain types of communicative behaviors—like grounding and responding to questions—are believed to facilitate successful communication (e.g., Clark & Brennan, 1991; White, 1997), perhaps by contributing to content and context during communication. Therefore, we were interested in the way these behaviors and current communicative success influenced lexical density. Our second set of analyses targets how three variables influence the amount of content that interlocutors provide one another (operationalized as lexical density) in each utterance: grounding, responding to a question, and communication state (i.e., miscommunication or successful communication).

In collaborative problem-solving tasks, the act of grounding usually refers to occasions in which an interlocutor confirms (e.g., through explicit verbal affirmation) a conversational partner's referent to an object in their shared environment. This process serves to increase an interlocutor's ability to find common ground by establishing shared knowledge in the current task. While grounding can often occur within the context of responding to a question, grounding and question-responding are distinct: A person can exhibit grounding behavior in response to their partner's statement (rather than a question), and they can respond to a question without grounding (e.g., asking another question, negating new information, providing a clarification rather than a new piece of information).

Specifically, individuals should tend to use more lexically dense language when engaging in grounding behaviors and when responding to a question, with a stronger association seen in successful communication (as opposed to miscommunication). During moments of grounding and when responding to a question, lexical density may increase as interlocutors try to establish novel referents or re-ground. However, when conversation is lexically shallow, interlocutors might not have the necessary information to communicate successfully.

Exploratory Analyses. We will also engage in exploratory analyses to better understand our findings and suggest new avenues of research into the impact of miscommunication. After conducting our planned analyses, we will conduct exploratory analyses to help better understand the effects observed. Because these will be exploratory (rather than *a priori*) analyses, these analyses will be guided by the specific results of the planned analyses.

Method

Participants

Participants included 20 dyads of paid undergraduate students from the University of Rochester who did not know one another before participating ($N = 40$; females[1] = 26; males = 14; mean age = 19 years). Participants were recruited through the university subject pool. All provided informed consent using IRB-approved procedures. All were native talkers of American English with normal to normal-corrected vision. None reported speech or hearing impairments.

[1] The experiment was run in 2012 and asked participants to self-report their gender using only "male" and "female" options, which are now associated with sex rather than gender.

Stimuli and Procedure

The current project analyzed a subset of a larger corpus aimed at capturing the linguistic and behavioral dynamics of dyadic task performance with and without shared visual fields (Paxton, Roche, Ibarra, & Tanenhaus, 2014; Paxton, Roche, & Tanenhaus, 2015; Roche, Paxton, Ibarra, & Tanenhaus, 2013; see similar paradigm in Ibarra & Tanenhaus, 2016).[2] Here, we analyzed the behavioral dynamics of only the interactions in which participants did not have a shared visual field. Participants engaged in a turn-taking task that required them to build a three-dimensional puzzle based on pictorial instruction cards. Participants were unable to see their partner, their partner's workspace, and their partner's instruction cards during the interaction; dyads coordinated building exclusively through spoken language exchanges. Interactions were transcribed and annotated for linguistic and behavioral measures.

Each data collection session was run by a single researcher[3], sometimes accompanied by an undergraduate research assistant who was blind to study hypotheses. Stimuli were two (2) bloco[tm] objects (www.blocotoys.com). Bloco objects are three-dimensional animal puzzles consisting of approximately 27 unique pieces each (grasshopper = 25 pieces; lizard = 28 pieces; see Fig. 1). During the condition analyzed here, each dyad was randomly assigned to construct only one of these two puzzles.

[Insert Figure 1 around here]

The building process was divided into an *Item phase* and a *Build phase* (see Table 1). During the *Item phase*, participants were asked to separate the individual building components anywhere within four square regions drawn on each participant's workspace. The participants

---

[2] The remainder of the corpus asked participants to engage in a similar task but asked participants to work together on the same object in a shared visual environment. Because of our operationalization of miscommunication (see "Measures" section below), this additional condition was not suitable for the current analyses.

[3] This researcher was either author J.R. or author A.I.

could freely decide together how to arrange the pieces, subject to two constraints: (1) Both participants needed to agree about where each of the objects should be placed; and (2) participants' separate workspaces must match one another's by the end of this phase. The *Item phase* facilitated participants' familiarity with each piece prior to the *Build phase* and tidied the workspace for easier building in the subsequent phase.

For the *Build phase*, we constructed a set of pictorial instruction cards that guided both participants through each step of the object-building process (see Figure 1B). The grasshopper puzzle required 13 steps, and the lizard puzzle required 15 steps. Each card displayed a single step and depicted only the pieces of the puzzle that were directly relevant to the current step. The cards were divided as evenly as possible between the participants (i.e., 8 versus 7 for the grasshopper puzzle and 7 versus 6 cards for the lizard puzzle).

After the *Item phase* was complete, participants were given the cards and were asked to work together to build the figure using the instruction cards. Although they were instructed to take turns providing the instructions, both participants could otherwise speak freely. Once they completed the final instruction, the experimenter informed the dyad whether they had correctly built the object. Two (2) dyads made minor mistakes after completing the figure (e.g., the grasshopper legs were upside-down). The pairs that did not construct the figure completely correctly were informed that something did not match and that they needed to identify and fix the errors (which all eventually did).

During the experiment, each dyad was video-recorded from three angles in order to obtain full views of each participant's workspace and to capture each participant in profile. This aided in coding the non-linguistic behavioral data through the course of the interaction (see "Measures" section below). The video recordings also captured audio, from which we fully transcribed the verbal exchanges between participants.

Open Code and Data

Due to assurances of confidentiality of data given to participants in the informed consent documents, we are unable to openly share the data for the project. The data were collected in 2012, prior to the widespread discussion of data-sharing that has since emerged in psychology and beyond. However, we have openly provided our code for analysis in our GitHub repository for our project: https://github.com/a-paxton/miscommunication-in-joint-action.

Measures

We transcribed each dyad's utterances along with several other non-linguistic behavioral measures. All transcription and coding procedures were performed by individuals who were blind to study hypotheses.

Turns. Using the audio data, a turn was coded as soon as one of the participants began to speak. When participants talked over one another, we maintained the turn structure by transcribing the talker who was "holding the floor" first and transcribing the talker who was "intruding" second. Across all 20 dyads, the corpus included a total of 8,493 turns.

Workspace Matching. In the present analyses, we quantify task success as the matching (or visual congruence) of partners' workspaces. An undergraduate research assistant (RA) coded the dyads' workspaces as either matching or mismatching on a turn-by-turn basis by examining the video streams for each dyad. The RA coded the visual environment at the end of each turn, the point at which one participant finished talking and before their partner began talking.

Often, a talker (T_a) was required to describe a spatial orientation to their partner (T_b). If T_b physically moved the object to the correct orientation (as intended by T_a based on by T_a's workspace and instruction card), the current turn was coded as having matching workspaces. However, if T_b failed to put the object in the correct orientation, the turn was coded as having mismatching workspaces. Figure 1C provides an imagined example of what a mismatched turn might look like. In this turn, T_a instructed T_b to orient the holes in an upward fashion, but the ambiguous use of "up" resulted in a visually incongruent turn—because the spatial term was applied to the referent in a way that was not intended by the talker.

Approximately 65% of the turns in the current subset of the corpus were successful communication turns (i.e., turns at the end of which participants' workspaces matched), while approximately 35% of the corpus were characterized by communication failure (i.e., turns at the end of which participants' workspaces mismatched). Thus, we were successful in creating a situation in which interlocutors communicated successfully with one another on most trials, yet local miscommunication occurred frequently enough to create a rich enough corpus for analysis.

We determined the coding reliability by having two additional hypothesis-blind coders with no prior knowledge of the experiment evaluate 5% of the visual congruence codes (425 turns) from the original RA codes. These coders were asked to determine whether they agreed or disagreed with the first RA's visual congruence codes for each turn. An inter-rater reliability analysis of these codes found high agreement with the primary coder (kappa = .96).

Lexical Density. We operationalize the amount of content in language as *lexical density*—that is, the ratio of content words to all words in a given utterance. We chose this over *lexical diversity* (i.e., another measure of language complexity that counts the total number of unique words in an utterance; cf. Johansson, 2008) because language can include a high level of lexical diversity (i.e., with many unique words) while still containing low lexical density (e.g., with many of the unique words being pronouns and auxiliaries instead of nouns and verbs; Bradac, Desmond,

& Murdock, 1977; Halliday, 1985; Johansson, 2008). Moreover, lexical density—as a ratio—naturally controls for the length of an utterance.

For our purposes, "content words" are nouns and verbs, excluding auxiliary verbs, pronouns, and very common words. The stopword corpus (i.e., a list of the most common words in a language, routinely removed from natural language processing because of their lack of situational specificity; e.g., pronouns, articles) in the `nltk` toolkit in Python formed the basis of our stopword list (Bird, Klein, & Loper, 2009). However, we removed from this list any of the lexical items of specific interest to our analyses (specified in the "Lexical Items" subsections below). A list of all stopwords in our analyses are included in our supplemental material.

Lexical density is a proportion of content words to total words. For example, if the words "green Christmas tree" comprised an entire turn, the turn would have a lexical density of 1, with 3 content words out of 3 total words. However, if the turn were "the green Christmas tree," it would contain 3 content words out of 4 total words, for a lexical density of 0.75.

Lexical Items: Assent and Negation. To facilitate automatic analysis, RAs transcribed the assent (e.g., *yes*, *yeah*, *yup*) and negation words (e.g., *no*, *nope*) using consistent spelling based on participants' utterances. Turns were then automatically annotated with separate binary variables for whether they included indications of assent and negation (0 = no words of that type included in the turn; 1 = at least 1 word of that type included in the turn). Assent and negation were not mutually exclusive—that is, a turn could be coded as 1 in assent and 1 in negation if that turn included at least one assent word and at least one negation word. A list of all identified assent and negation terms in our analyses and the software code used to implement the automatic annotation are included in our supplemental material on GitHub.

Lexical Items: Spatial Terms. We identified spatial terms (e.g., *up*, *down*, *left*, *right*)—which are likely to be ambiguous in the current task because of the lack of shared visual information—by examining the unique words uttered by all participants to find words that could be spatial in nature. We then confirmed that these words were used as spatial markers by reading through the turns in which these identified terms occurred. Potential words that were not used as spatial referents in the majority of turns were not considered to be spatial terms. As with assent and negation, turns were then automatically annotated with a binary variable for whether they included a spatial term (0 = no spatial words; 1 = at least 1 spatial word). A list of all identified spatial terms in our analyses and the software code used to implement the automatic annotation are included in our supplemental material on GitHub.

Pragmatic Behavior: Grounding. Grounding was manually coded by two coders (author J.R. and A.I.) using a procedure similar to the one described by Nakatani and Traum (1999). Grounding was established through evaluating *grounding units*, in which one talker presented a new piece of information. A turn was marked as grounded when the unit was accepted by the other talker (in Fig. 1C, T_a: *Do you want to put, like, all the green ones in that box, or…?*; T_b: *Okay*.). The coders reached 87.5% agreement and substantial inter-rater reliability ($\kappa = .61$; see Landis & Koch, 1977). For instances that agreement was not met in the initial ratings, the two coders discussed the discrepancies until consensus on the code was reached.

In the current analyses, we only counted explicit verbal grounding (i.e., at least one verbal indication in the turn immediately following one in which their partner offered new information). This did not have to be explicit assent but could include any kind of acknowledgement or response to their partner (e.g., responding with a location or direction).

Pragmatic Behavior: Response to Questions. Utterances containing an implicit or explicit question were indicated by the RA in the transcription with a question mark; these turns were counted as including questions. The utterance immediately following that turn (which was necessarily their partner's turn in the present transcription scheme) was automatically marked with our software as being a response to question. For instance, if one member of the dyad (T_a) asked a question (as marked by a question mark in the transcription), the other member of the dyad (T_b) would be marked as "responding to a question" in the next turn. Turns marked as being a response to a question were not *necessarily* marked as grounding, although they *could* also be marked as grounding if grounding verbal behavior occurred during the response (see previous description). This relatively crude measure—again, simply marking whether the turn was preceded by one in which a question was asked by their partner—allowed us to capture information about question-responding behavior.

Analytic Approach

All analyses were performed in R (R Development Core Team, 2012), with all models built using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015). Each model reported below includes the maximal random effect structure supported by the data with dyad identity and turn number set as random intercepts. Each intercept included the maximal random slope structure justified by the data (using backward selection or "leave-one-out-method" until reaching convergence; Barr, Levy, Scheepers, & Tily, 2013). For clarity and ease of reading, we present all model results in tables and refer to the specific predictors in the text.

All dichotomous variables were dummy-coded and centered: whether the turn ended in miscommunication (-0.5 = matching state; 0.5 = mismatching state), whether grounding occurred during the turn (-0.5 = not grounded; 0.5 = grounded), whether the turn did not include (-0.5) or included (0.5) at least one word from our target lexical items (assent, negation, and spatial words), and whether the turn was a response to a question (-0.5 = not a response to a question; 0.5 = response to a question). All main effects and interaction terms were centered and scaled prior to entry into the model, permitting estimates to be interpreted as effect sizes (Keith, 2005).

As discussed in the Method section, lexical density was calculated by dividing the number of content words by the number of total words in a turn, creating a natural floor and ceiling for the variable). After inspecting the data, we observed that participants used a number of one-word utterances (e.g., "*Yeah*," "*No*," "*Up*") over the course of the task, creating a large number of turns at the ceiling or floor of lexical density. This means that it could be difficult to determine whether greater lexical density is having an effect (i.e., over the whole range of possible lexical density values; as we hypothesized) versus whether any effect of lexical density is driven by two additional possibilities: by one-word turns (i.e., which could only be at ceiling or at floor) or by turns with maximum lexical density (i.e., hitting the ceiling of the lexical density value). To rule out the possibility that our results were artifacts of the ceiling of lexical density or the presence of one-word turns, Models 1 and 2 were each constructed using multiple subsets of the data: (A) the full dataset (total turns = 8,494), (B) excluding MLD turns (i.e., turns with maximum lexical density; included turns = 3,341), and (C) excluding turns comprising only one word, which we call *OW turns* (included turns = 2,278). All unstandardized models are available at the GitHub repository for the project (see above).

Model 1. Model 1 evaluated the effects of pragmatic and lexical items (spatial, assent, negation, response to question, and lexical density) on successful communication (matching) and miscommunication (mismatching) turns using mixed-effects logistic regressions.

Model 2. To answer this question, we analyzed lexical density by grounding, responding to questions, and communicative state (along with their interactions) using linear mixed-effects models for three datasets: full turns, without MLD turns, and without OW turns. Moreover, exploring the patterns of lexical density may help shed light on some of the effects in Model 1.

Exploratory Analyses. Exploratory analyses will be conducted to investigate interesting patterns observed in Models 1 and 2. However, because they are contingent on the results from our planned models, we did not approach the exploratory analyses with a specific analysis plan in mind.

Results

Model 1

Model 1A: Full Data (Table 2). As hypothesized, successful communication was more likely to be associated with higher lexical density and the presence of assent words and that miscommunication was more likely to be associated with the use of spatial terminology (i.e., ambiguous language). As anticipated, we also saw a trend toward a positive relation between negation word use and miscommunication, although it did not reach statistical significance. Contrary to our hypothesis, however, we found that responses to a question were more likely to be associated with miscommunication at the end of the turn.

[Insert Table 2 around here]

Model 1B: Without Maximum Lexical Density (MLD) Turns (Table 3). Results were nearly identical to the raw model, with the exception that lexical density no longer predicted communication state but trended in a similar direction. Differences between the models with and without MLD turns could be driven by one-word turns (i.e., producing ceiling or floor effects).

[Insert Table 3 around here]

Model 1C: Without One-Word (OW) Turns (Table 4). Results were identical to the patterns found in our analysis of MLD turns (Model 1B): Negation again trended toward an effect but did not reach significance, and lexical density again failed to significantly predict communication state. Although we cannot conclusively discriminate between the effects of OW and MLD turns, these results suggest that OW/MLD turns drove the effect of lexical density observed in the full dataset but that the other effects were robust across all turns.

[Insert Table 4 around here]

Model 2

Model 2A: Full Data (Table 5). As expected, greater lexical density was positively associated with grounding. Contrary to expectations, however, lexical density was negatively connected with responding to a question, such that interlocutors tend to use shallower language when answering a partner's question. We found a trend toward dyads using lexically shallow turns during miscommunication, although it did not reach statistical significance.

Against our expectations, we did not find that successful communication amplified the effects of grounding and responding to a question. However, dyads tended to produce more lexically shallow language when participants were grounding and responding to a question simultaneously (see Fig. 2): When asked a question that offered a new piece of information or re-established a lexical pact, the interlocutor's response tended to be less content-full. Interestingly, dyads were most lexically dense when grounding in response to statements (not questions). This could indicate verbal tracking or OW assent turns (e.g., saying "Uh-huh" in response to a partner's statement to imply understanding).

[Insert Table 5 around here]

[Insert Figure 2 around here]

[Insert Figure 3 around here]

Model 2B: Without MLD Turns (Table 6, Fig. 3). Results were nearly identical to Model 2A, with two exceptions: Mismatch state no longer trended toward significance, and the interaction between grounding behavior and responding to a question no longer reached significance, although it trended in a similar direction. These were again congruent with the possibility that OW assent turns—which would be marked as MLD—drove these effects. Our next model then tests whether removal of OW turns shows similar effects.

[Insert Table 6 around here]

Model 2C: Without OW Turns (Table 7, Fig. 3). Results were identical to Model 2A, supporting our intuition that these effects could be largely driven by OW assent turns.

[Insert Table 7 around here]

Exploratory Analysis (Model 3, Table 8)

As noted in our Analytic Approach section, we used our results from Models 1 and 2 to guide our choices in our exploratory analysis in Model 3. OW and MLD turns appeared to drive a number of effects in Model 2, but the invariance of lexical density in both subsets of the data leave us unable to disentangle these possible effects according to the amount of content being shared between talkers. Because Models 2C and 2B would both remove turns that included a single assent word (e.g., "yeah" or "uh-huh"), neither Model 2B nor Model 2C would be able to capture back-channeling. We identified OW assents as a potential means of disentangling the contributors to miscommunication in OW and MLD turns. When participants respond to one another with a single assent word, miscommunication could arise if the talker intends the assent to be a form of verbal tracking (or back-channeling) while the listener interprets it as grounding (e.g., saying "uh-huh" to affirm attention, not understanding). Therefore, we used our exploratory model to evaluate assent words in a dataset that only included maximally dense utterances, using grounding, response to a question, mismatch state, and all permissible interactions[4] as predictors. To do so, we created a fourth (and final) dataset that *included* only maximally dense turns (turns = 5,460).

Our exploratory model found a significant main effect of grounding and response to a question and a significant interaction between grounding and mismatch state. Consistent with previous literature, dyads were significantly more likely to use an assent word when grounding. (Again, grounding did not necessarily have to include an assent word; any explicit acknowledgement or building onto a previous statement would be considered grounding.)

Interestingly, dyads were *less* likely to use an assent word when responding to a question with an MLD turn, suggesting that participants tended to spend more time and (lexical) effort when responding to one another's inquiries. Although responding with only a "Yes" or "No" would be perfectly lexically dense, interlocutors did not necessarily do that. Instead, the dyads appeared to provide "bite-sized" information that could be more targeted than a simple affirmation. When grounding, dyads were equally likely to assent during successful and miscommunication turns; when not grounding, they were more likely to assent during successful communication (see Fig. 4).

[Insert Table 8 around here]

[Insert Figure 4 around here]

Discussion

Miscommunication arises regularly during interaction in everyday life—especially in the context of joint action or shared goals. Our current corpus reflects this reality, with

[4] Only the interaction between grounding and mismatch state could be included in this analysis. All other interactions did not include sufficient observations over the possible combinations to achieve convergence.

miscommunications occurring in approximately 35% of communicative turns in a collaborative dyadic task that asked participants to bridge distributed instructions to build puzzle objects without being able to see one another or one another's workspaces. As in everyday life, interlocutors were able to successfully complete a cognitively complex but mechanically simple task together despite ample miscommunication. We examine the effects of pragmatic and lexical behaviors on miscommunication, building on previous work on communicative processes that lead to successful communication and exploring how they function in miscommunication.

Pragmatic and Lexical Predictors of Miscommunication

Our first analysis unpacked the language dynamics associated with moment-to-moment miscommunication (Model 1A). Some behaviors—when an interlocutor was answering a partner's question or using more ambiguous task-specific language (i.e., spatial terms)—were more likely to result in miscommunication. Spatial terminology was particularly problematic because the dyads lacked a shared visual space during an inherently spatial task, although the interlocutors were still successfully able to use spatial terminology at least half of the time. While our task may appear somewhat unnatural, our connected societies are increasingly supporting remote collaboration—including during contexts without shared visual fields. The key to success is ensuring that ambiguity is grounded in relation to the current referent and within the current communicative context. Failure to appropriately ground appears to be the primary link between communication breakdown and spatial terminology.

We also saw a trend toward negation language leading to miscommunication, although it failed to reach statistical significance. Other behaviors—like using more assent words or more lexically dense language—were associated with successful communication. This is consistent with previous literature finding that interlocutors' production strategies often facilitate communication (e.g., grounding, Bazzanella & Damiano, 1999; Clark & Brennan, 1991). Agreement's association with success is perhaps unsurprising, but it does lend support to the intuitive idea that partners use assent meaningfully and not simply as filler or backchanneling. Follow-up analyses controlling for maximal lexical density (Model 1B) and minimal turn length (Model 1C) found these results to be quite robust: Turns that included a question or more task-specific ambiguous language were consistently more likely to end in a state of miscommunication, while turns that included an indication of assent were consistently more likely to end in a state of successful communication.

Interactive collaborative conversation requires a balance of task success with language production costs. One way in which interlocutors reduce cognitive effort is by limiting the amount of explicit information in their utterances (Levinson, 1983)—including by relying on their context and environment to disambiguate (Piantadosi et al., 2012). If interlocutors have fully established referents, ambiguous language can help reduce redundancy and processing load (Aylett & Turk, 2004; Levy & Jaeger, 2007; Piantadosi et al., 2012). However, ambiguous language can become problematic if the context is not sufficiently rich or if referents are not appropriately established.

We also evaluated contexts in which lexically shallow utterances have the potential to hurt communication, keeping in mind that lexically shallow utterances might be more ambiguous than lexically dense utterances. Miscommunication was associated more with lexically shallow utterances than was successful communication. Lexical density—that is, using a higher percentage of "content-full" words (like nouns and verbs) per turn (rather than, e.g., pronouns or articles)—is closely tied to Gricean maxims, especially the idea that talkers should provide precisely and only the amount of information needed by the listener (Grice, 1975). Lexical density was linked to successful communication in longer turns but this effect did not hold when controlling for maximum lexical density and single-word turns. These findings support the idea that variability of content may play a key role in successful communication: Partners work together smoothly when they include more content per turn but not when the turn is completely saturated (Grice, 1975).

However, we cannot always know what our conversational partner knows or is currently experiencing. This makes communication difficult. In fact, lexically dense utterances are more often associated with successful communication in the full dataset (Model 1A), suggesting that the investment of effort can lead to improvement. This is consistent with complementary findings from previous research that finds that talkers are more likely to be over- rather than under-informative, even linking more successful communication to more lexically dense communication (Davies & Katsos, 2010; Engelhardt, Bailey, & Ferreira, 2006; Pogue, Kurumada, & Tanenhaus, 2016). A notable exception, however, is use of referring expressions in task-based practical dialogues where dyads engage in extended dialog. Under these circumstances, under-modification is extremely common (Brown-Schmidt & Tanenhaus, 2008).

Despite these similarities to previous research, our results suggest some nuance when we try to parse the effects of lexical density. Our follow-up models (Models 1B and 1C) found some evidence that the effect of informativeness is driven by extremely short and/or extremely dense turns, suggesting an avenue for future research.

Contributors to Lexical Density during Collaborative Task Performance

When analyzing the entire dataset (Model 2A), we found that lexical density increased with grounding. However, when interlocutors responded to a question with grounding or in a state of miscommunication, their utterances were typically lexically shallow. Dyads were least lexically dense when responding to a question without grounding and most lexically dense when responding to statements while grounding.

Although lexically shallow utterances could lead to miscommunication through under-specification, reducing lexical richness could facilitate long-term communicative success by prompting interlocutors to “check back in” with one another. Miscommunication may boost the integrity of the communication system by helping facilitate deeper understanding when required but otherwise allowing us to conserve cognitive resources (Haywood et al., 2005; Horton & Keysar, 1996; Roche, Dale, & Kreuz, 2010). Miscommunication may bootstrap a general cognitive process (e.g., monitoring and adjustment; Horton & Keysar, 1996) that encourages an investment of cognitive effort only when the context demands it and provides *cheap* and *simple* strategies to resolve miscommunication (see Svennevig, 2008).

These patterns were stable even when controlling for very lexically dense turns (Model 2B), with the notable exception that the interaction between grounding and response to questions was no longer significant. Follow-up analyses further suggested that—in longer utterances—interlocutors tend to be more lexically dense when grounding but tend to use shallower language when responding to a question (Model 2C). Our ability to disentangle the possible effects of very short and very dense language, however, was limited due to the restricted variability of lexical density across the two subsets. This pushed us to look outside of the effects of lexical density and to indications of assent: It could be that turns comprising only assent words could lead to different patterns of success, depending on how they are used.

Because assent words have the potential to indicate understanding or attention, our final model (Model 3) evaluated whether the presence of an assent could differentially predict miscommunication in maximally lexically dense turns. Previous work has found that interlocutors tend to use assent as an affirmation of understanding or for affirmation of attention (Bavelas & Gerwing, 2011; Lambertz, 2011; Yngve, 1970). Congruent with previous work, we found that assent words acted both as a way to ground during smooth communication and as a way to positively affirm one’s attention to the current context in the face of miscommunication.

This “multitasking”—the context-sensitive meaning of assent terms given the situation—may be a significant contributor to miscommunication: A listener may misinterpret an assent as an affirmation of understanding when it was meant as an affirmation of attention (or vice-versa). We find that the processes underlying successful communication are also present during miscommunication—but their context-sensitivity leads them to function differently, leading to different outcomes.

Limitations and Future Directions

Here, we have only considered spatial terminology as a type of ambiguous language and did not include other forms of ambiguous communication (e.g., omission). This task was designed for unscripted language use, which benefits by capturing natural language patterns but may result in a loss of experimental control. In addition, the complexity of language and interaction likely means that a host of other pragmatic and lexical factors (outside of the scope of the current paper) also affected the conversation context and task performance.

However, the naturalistic nature of the task allowed us to contribute to the growing body of work on joint action and communication, supporting the idea that miscommunication may help bring greater attention to bear on the situation during difficult moments in interaction. This task also provides insights that may be used to design more targeted language-game experiments to explore the effects of pragmatic and lexical behaviors on communicative success and failures.

Though our current study does not speak directly to learning, our findings lead us to question more deeply what role miscommunication has on the communicative system. Future work should explore how miscommunication affects higher levels of socio-pragmatic effects on communication, like rapport. This may be done by evaluating behavioral alignment (cf. Paxton et al., 2014) and self-reports of perceived rapport. Future work should also look at learning gains that may occur during moments of uncertainty and ambiguity resolution: Miscommunication's perturbation of the system could require the user to invest more effort cognitively, increasing the likelihood of encoding information into long-term memory.

Implications

Our findings—while basic research about low-stakes miscommunication contexts—have implications for high-pressure contexts, like the medical contexts we discussed in the opening of the paper (e.g., Halverson et al., 2011; Isaacs & Creinin, 2003; Lingard et al., 2004; Phillips et al., 2001; Raley et al., 2016; Sutcliffe, Lewton, & Rosenthal, 2004). Our results support a view of miscommunication as highly efficient for cognitive load, reducing individual strain by offloading it to the dyadic system: Rather than constantly investing precious cognitive resources in over-specifying information, interlocutors wait for the context (most notably, their partner) to nudge them into investing effort only when necessary. Waiting for these nudges is relatively benign in the current experimental context; failure only means waiting a bit longer before leaving the experiment. Clearly, such a strategy is untenable for medical contexts with life-or-death consequences or other high-stakes situations.

However, our findings dovetail with a growing literature on reducing workplace accidents and malpractice that relies not on individuals maintaining constant (and taxing) vigilance but on a *system* that will offload some of that cognitive strain (e.g., Harry & Sweller, 2016), including other people (e.g., Young, ten Cate, O'Sullivan, & Irby, 2016). Cognitive aids—tools like checklists and manuals—improve patient outcomes by accounting for cognitive load among the caregiving team (e.g., Fletcher & Bedwell, 2014; Goldhaber-Fiebert & Howard, 2013) in the face of the view of (mis-)communication and (under-)specification demonstrated here in joint action contexts. Acknowledging that these high-stakes contexts are an outgrowth of normal human communicative processes and continuing to elucidate those dynamics through basic research will be critical to reducing miscommunication during life-or-death settings as well as more contrived ones.

Conclusion

Using language to facilitate joint action requires interlocutors to maintain a constant balance of effort between listeners and talkers, and we find that miscommunication may help the dyadic system achieve that balance. Brief communicative "stumbles" may help us communicate more effectively within our contextual and physical constraints, pushing us to check back in with one another, help us re-establish mutual understanding, and push us to further ground our interaction. Miscommunication may both emerge *and* benefit from the cost-saving cognitive processes associated with shallow and ambiguous language. As such, we point to the importance of miscommunication and its ramifications—suggesting, perhaps, that miscommunication may be as critical to interaction as successful communication.

Acknowledgements

Special thanks go to our undergraduate research assistants at University of Rochester (Chelsea Marsh, Eric Bigelow, Derek Murphy, Melanie Graber, Anthony Germani, Olga Nikolayeva, and Madeleine Salisbury) and University of California, Merced (Chelsea Coe and J.P. Gonzales). We would also like to thank the organizers (Patrick Healey and J.P. de Ruiter) and attendees of the 3rd International Workshop on Miscommunication for their feedback and suggestions for this project. Author Alyssa Ibarra is now affiliated with the Office of Institutional Research at the University of Rochester.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., … Weinert, R. (1991). The Hcrc Map Task Corpus. *Language and Speech*. https://doi.org/10.1177/002383099103400404

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. https://doi.org/10.1177/00238309040470010201

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*. https://doi.org/10.1080/10904018.2010.508675

Bazzanella, C., & Damiano, R. (1999). The interactional handling of misunderstanding in everyday conversations. *Journal of Pragmatics*. https://doi.org/10.1016/S0378-2166(98)00058-7

Bird, S., Klein, E., & Loper, E. (2009). Language Processing and Python. *Computing*.

Bradac, J. J., Desmond, R. J., & Murdock, J. I. (1977). Diversity and density: Lexically determined evaluative and informational consequences of linguistic complexity. *Communication Monographs*, *44*(4), 273–283. https://doi.org/10.1080/03637757709390139

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation.

Journal of Experimental Psychology: Learning Memory and Cognition. https://doi.org/10.1037/0278-7393.22.6.1482

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*. https://doi.org/10.1080/03640210802066816

Campana, E., Tanenhaus, M. K., Allen, J. F., & Remington, R. (2011). Natural discourse reference generation reduces cognitive load in spoken systems. *Natural Language Engineering*. https://doi.org/10.1017/S1351324910000227

Cardosi, K. (1998). Human factors lessons learned in the design and implementation of air traffic control systems. *The Controller*, *1*, 11–15.

Clark, H., & Brennan, S. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition.*, *13*, 127–149.

Clark, H. H., & Carlson, T. B. (1982). Speech acts and hearers' beliefs. In *Mutual Knowledge*.

Clark, H. H., & Marshall, C. R. (1981). *Definite reference and mutual knowledge*. (A. K. Koshi, B. Webber, & I. A. Sag, Eds.), *Elements of Understanding*. Cambridge: Cambridge University Press.

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

D'Mello, S. K., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, *25*(7), 1299–308. https://doi.org/10.1080/02699931.2011.613668

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*.

https://doi.org/10.1016/j.lingua.2010.02.005

Eisenberg, E. M., Murphy, A. G., Sutcliffe, K., Wears, R., Schenkel, S., Perry, S., & Vanderhoef, M. (2005). Communication in Emergency Medicine: Implications for Patient Safety. *Communication Monographs*, *72*(4), 390–413. https://doi.org/10.1080/03637750500322602

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2005.12.009

Fletcher, K. A., & Bedwell, W. L. (2014). Cognitive aids: Design suggestions for the medical field. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *3*(1), 148-152. https://doi.org/10.1177/2327857914031024

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*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2003.10.016

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*. https://doi.org/10.1207/s15516709cog1703_1

Goldhaber-Fiebert, S. N., & Howard, S. K. (2013). Implementing emergency manuals: Can cognitive aids help translate best practices for patient care during acute events? *Anesthesia & Analgesia*, *117*(5), 1149-1161. https://doi.org/ 10.1213/ANE.0b013e318298867a

Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, *95*(3), 524–536. https://doi.org/10.1037/0022-0663.95.3.524

Grice, P. (1975). Logic and conversation. In R. Stainton (Ed.), *Perspectives in the philosophy of language: A concise anthology* (pp. 41–58). Broadview.

Guerrero, L. K., Andersen, P. A., & Afifi, W. A. (2001). *Close Encounters: Communicating in Relationships.* Mountain View, CA: Mayfield.

Halliday, M. A. K. (1985). *Spoken and written language*. Geelong VIC: Deakin University.

Halverson, A. L., Casey, J. T., Andersson, J., Anderson, K., Park, C., Rademaker, A. W., & Moorman, D. (2011). Communication failure in the operating room. *Surgery*, *149*(3), 305–310. https://doi.org/10.1016/j.surg.2010.07.051

Harry, E., & Sweller, H. (2016). Cognitive load theory and patient safety. In K. J. Ruskin, S. H. Rosenbaum, and M. P. Stiegler (Eds.), *Quality and safety in anesthesia and perioperative care*. New York: Oxford University Press.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, *16*(5), 362–366. https://doi.org/10.1111/j.0956-7976.2005.01541.x

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Ibarra, A., & Tanenhaus, M. K. (2016). The flexibility of conceptual pacts: Referring expressions dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology*, *7*(APR). https://doi.org/10.3389/fpsyg.2016.00561

Isaacs, J. N., & Creinin, M. D. (2003). Miscommunication between healthcare providers and patients may result in unplanned pregnancies. *Contraception*, *68*(5), 373–376. https://doi.org/10.1016/j.contraception.2003.08.012

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers in Linguistics*, *53*, 61–79.

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language Problems & Language Planning*, *27*, 233–248. https://doi.org/10.1075/lplp.27.3.03jon

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*. https://doi.org/10.1515/IP.2007.004

Lambertz, K. (2011). Back-channelling: The use of yeah and mm to portray engaged listenership. *Griffith Working Papers in Pragmatics and Intercultural Communication*, *4*(1/2), 11–18.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, *2*(2), 205–217. https://doi.org/10.1093/semant/2.2.205

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *NIPS*. https://doi.org/10.1111/j.0956-7976.2005.01541.x

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., … Grober, E. (2004). Communication failures in the operating room: an observational classification of recurrent types and effects. *Quality & Safety In Health Care*, *13*(5), 330–4. https://doi.org/10.1136/qhc.13.5.330

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24. https://doi.org/10.1111/j.1467-9280.2008.02041.x

McTear, M. (1991). Handling miscommunication in spoken dialogue systems: why bother? In K. Mogford-Bevan & J. Sadler (Eds.), *Child Language Disability* (2nd ed., pp. 19–42). Clevedon: Multilingual Matters.

McTear, M. (2008). *Handling miscommunication: Why bother?* (Recent tre). Dordrecht: Springer.

Mills, G. (2014). Establishing a communication system: Miscommunication drives abstraction. In *Evolution of Language: Proceedings of the 10th International Conference* (pp. 193–194).

Nakatani, C. H., & Traum, D. R. (1999). *Coding discourse structure in dialogue (version 1.0)*. Technical Report UMIACS-TR-99-03.

Paxton, A., Roche, J. M., Ibarra, A., & Tanenhaus, M. K. (2014). Failure to (mis) communicate: Linguistic convergence, lexical choice, and communicative success in dyadic problem solving. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Paxton, A., Roche, J. M., & Tanenhaus, M. K. (2015). Communicative efficiency and miscommunication: The costs and benefits of variable language production. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society.* Austin, TX.

Phillips, J., Beam, S., Brinker, A., Holquist, C., Honig, P., Lee, L. Y., & Pamer, C. (2001). Retrospective analysis of mortalities associated with medication errors. *American Journal of Health-System Pharmacy*, *58*(19), 1835–1841. https://doi.org/ 10.1093/ajhp/58.19.1835

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.02035

R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. *R Foundation for Statistical Computing, Vienna, Austria.*

Raley, J., Meenakshi, R., Dent, D., Willis, R., Lawson, K., & Duzinski, S. (2016). The Role of Communication During Trauma Activations: Investigating the Need for Team and Leader Communication Training. *Journal of Surgical Education*, 1–7. https://doi.org/10.1016/j.jsurg.2016.06.001

Roche, J. M., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during conversation: More than mere mimicry? *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 206–211. Retrieved from https://www.hlp.rochester.edu/resources/workshop_materials/EVELIN12/RocheETAL10_disambiguation.pdf

Roche, J. M., Paxton, A., Ibarra, A., & Tanenhaus, M. K. (2013). From minor mishap to major catastrophe: Lexical choice in miscommunication. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 3303–3308. Retrieved from http://mindmodeling.org/cogsci2013/papers/0588/paper0588.pdf

Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of “uh huh” and other things that come between sentences. In *Analyzing discourse: text and talk*. https://doi.org/10.2307/324165

Sutcliffe, K. M., Lewton, E., & Rosenthal, M. M. (2004). Communication Failures: An Insidious Contributor to Medical Mishaps. *Academic Medicine*. https://doi.org/10.1097/00001888-

200402000-00019

Svennevig, J. (2008). Trying the easiest solution first in other-initiation of repair. *Journal of Pragmatics*. https://doi.org/10.1016/j.pragma.2007.11.007

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes*, *23*(3), 451–470. https://doi.org/10.1111/j.0883-2919.2004.00368.x

White, R. (1997). Back channelling, repair, pausing, and private speech. *Applied Linguistics*, *18*(3), 314–344. https://doi.org/10.1093/applin/18.3.314

Yngve, V. (1970). On getting a word in edgewise. *Papers from the 6th Regional Meeting of the Chicago Linguistic Society*, 568.

Young, J. Q., ten Cate, O., O’Sullivan, P. S., & Irby, D. M. (2016). Unpacking the complexity of patient handoffs through the lens of cognitive load theory. *Teaching and Learning in Medicine*, *28*(1), 88-96. https://doi.org/10.1080/10401334.2015.1107491

Zipf, G. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA.

List of Figure Captions

Figure 1. Panel A: Grasshopper (left) and lizard (right) Bloco figures used in the current study. Panel B: Sample instruction cards for the grasshopper figure (left) and lizard figure (right). Panel C: Example of Bloco items oriented differently that may lead to miscommunication; here, *up* is infelicitously indexed.

Figure 2. Lexical density when the response to a question (not answering - *left*; answering - *right*) was grounded (green) or not grounded (purple) in the full dataset (Model 2A). Bars represent standard error.

Figure 3. Lexical density when not grounding (left) or grounding (right) in response to a question during matching (blue) and mismatching workspaces (red) across the three datasets used in Models 2A, 2B, and 2C (from left to right: full data, without MLD turns, and without OW turns). Bars represent standard error.

Figure 4. Use of assent words when not grounding (left) or grounding (right) during mismatching workspaces (red) and matching (blue) workspaces. Bars represent standard error.

List of Tables and Table Captions

Table 8. Results of exploratory analysis predicting the use of assent words with grounding, response to a question, and workspace state during one-word turns (Model 3).

Table 1

Phase	Goal	Structure	Duration
Phase I: Item	Arrange all puzzle pieces for Bloco objects in identical patterns on their individual workspaces	No turn-taking instructions from experimenter; completely free conversation	mean time = 8.26 min mean turns = 14.38 turns
Phase II: Build	Assemble all puzzle pieces to create identical Bloco objects in their individual workspaces	Instruction cards divided in alternating order between both participants to create alternating instruction-givers; otherwise completely free conversation	mean time = 23.34 min mean turns = 19.07 turns

Table 2

Effect	*ß*	*SE*	*z*	*p*
Response to question	0.238	0.0624	3.823	$<.001^{***}$
Spatial word used	0.132	0.046	2.876	0.004^{**}
Assent word used	-0.133	0.027	-4.909	$<.001^{***}$
Negation word used	0.101	0.054	1.862	0.06.
Lexical density	-0.063	0.029	-2.14	0.03^{*}

Table 3

Effect	*ß*	*SE*	*z*	*p*
Response to question	0.240	0.064	3.747	<.001***
Spatial word used	0.146	0.061	2.389	0.02*
Assent word used	-0.105	0.031	-3.342	0.001**
Negation word used	0.113	0.059	1.899	0.06.
Lexical density	-0.045	0.031	-1.454	0.15

Table 4

	ß	*SE*	*z*	*p*
Response to question	0.097	0.029	3.295	0.001^{**}
Spatial word	0.134	0.053	2.509	0.01^{*}
Assent word	-0.132	0.031	-4.217	$<.001^{***}$
Negation word	0.109	0.061	1.789	0.07.
Lexical density	-0.039	0.031	-1.276	0.2

Table 5

Effect	*ß*	SE	*t*	*p*
Grounded	0.379	0.049	7.725	$<.001^{***}$
Response to question	-0.396	0.017	-23.450	$<.001^{***}$
Mismatch state	-0.075	0.042	-1.776	0.08.
Grounded x Mismatch state	0.017	0.020	0.867	0.39
Grounded x Response to question	-0.094	0.019	-4.882	$<.001^{***}$
Mismatch state x Response to question	0.029	0.020	1.453	0.15
Grounded x Mismatch state x Response to question	-0.019	0.020	-0.966	0.33

Table 6

Effect	*ß*	*SE*	*t*	*p*
Grounded	0.360	0.059	6.007	$<.001^{***}$
Responded to question	-0.081	0.023	-3.455	0.001^{**}
Mismatch state	-0.068	0.052	-1.305	0.19
Grounded x Mismatch state	-0.029	0.025	-1.188	0.23
Grounded x Response to question	-0.012	0.024	-0.517	0.61
Mismatch state x Responded to question	0.005	0.025	0.237	0.81
Grounded x Mismatch state x Responded to question	-0.014	0.025	-0.577	0.56

Table 7

	ß	*SE*	*t*	*p*
Grounded	0.325	0.052	6.236	$<.001^{***}$
Responded to question	-0.175	0.022	-7.815	$<.001^{***}$
Mismatch state	-0.055	0.050	-1.088	0.28
Grounded x Mismatch state	-0.008	0.023	-0.320	0.75
Grounded x Responded to question	-0.045	0.023	-1.937	0.05.
Mismatch state x Responded to question	0.005	0.025	0.196	0.84
Grounded x Mismatch state x Responded to question	-0.0154	0.0234	-0.647	0.52

Table 8

	ß	*SE*	*z*	*p*
Grounded	1.449	0.191	7.586	$<.001^{***}$
Responded to question	-0.378	0.047	-7.768	$<.001^{***}$
Mismatch state	-0.358	0.191	-1.874	0.06.
Grounded x Mismatch state	0.229	0.092	2.492	0.01^{*}