Collaboratively Setting Perspectives and Referring to Locations Across Multiple Contexts

Susan E. Brennan[1] (susan.brennan@stonybrook.edu)
Katharina S. Schuhmann[2] (katharina.schuhmann@stonybrook.edu)
Karla M. Batres[1] (karla.batres@stonybrook.edu)

Departments of [1]Psychology & [2]Linguistics,
Stony Brook University,
Stony Brook, NY 11794-2500 USA

Abstract

The Walking Around Corpus is an experimentally designed corpus of spontaneous spoken dialogs in which a stationary partner (a Giver, *G*) directed a pedestrian (a Follower, *F*) over a mobile phone to visit 18 destinations over a ~1.8 mile route on the Stony Brook University campus. After the navigation task, each of the 36 pairs participated in several traditional laboratory tasks, including individual memory tests for the locations, tests of spatial ability, self-reported experience with campus, and 6 rounds of a referential communication task in which pairs matched pictures of the target locations. We are examining how speakers adapt referring expressions across contexts, as well as the degree to which pairs entrain as they repeatedly referred to locations. The sequence of tasks enables us to examine how perspectives (and the lexical choices that index them) are shaped in both private and shared contexts. We have also examined effects of having visual context and of individuals' spatial ability upon performance. The Walking Around Corpus is a resource for studies of spoken vocabulary and spatial language, as well as for designing pedestrian GPS-based navigation applications. Over 30 hours of spoken dialogs are accompanied by detailed transcriptions that include disfluencies, pauses, and time stamps. The corpus is available for research purposes (contact the first author).

Keywords: Referential communication; spatial cognition; visual context; conceptual pacts; entrainment; collaboration; mobile communication; GPS apps; pedestrian navigation.

Variability in Spoken Dialog

Language production in spoken dialog is remarkably variable (see, e.g., Brennan & Clark, 1996); there are many ways in which to express the "same" message. Yet this variation is not random; individual speakers are constrained by memory and previous experiences, as well as by the communicative contexts in which they find themselves during everyday language use. Speakers are remarkably flexible in adapting to different contexts and conversational partners; their utterances are shaped not only by automatic or implicit processes but also by strategic or explicitly intention-driven processes. Although sometimes inferences slow processing, it has been shown that when contextual information is readily available—for instance, when an inference has already been made about a partner's communicative needs—a speaker may be able to use this information flexibly and without delay to adapt speaking for that partner (e.g., Galati & Brennan, 2010, 2013; Hwang, Brennan, & Huffman, submitted).

When speakers are physically co-present in the same dialog context (e.g., face-to-face in the same time and place), they may have abundant perceptual cues: they can see each other's reactions, non-verbal cues, and any task-related actions. When their utterances are mediated (e.g., by telephone, whether by voice or by text), they may represent the partner's presence, although the cues are diminished (e.g., with audible but not visual cues for grounding). When speakers use language in a solitary context (such as in a memory experiment, responding to to a computer prompt), they may design their utterances for an imagined or a presumed witness (e.g., an experimenter), or they may not think much about the audience at all (such as when musing to themselves).

The point is that there is no "default" context—only unknown or uncontrolled contexts (in the case of many language experiments). Language use is flexible enough that the context in which speakers find themselves is likely to have profound effects on the timing and form of utterances, such that it is risky for a psycholinguist or computational linguist to presume that ignoring context and modeling a default or generic situation necessarily probes basic or pure processes and representations. Many corpora exist that can be used to build and test theories about language. Understanding the ways in which these corpora were collected (the contexts of language use that these corpora represent and the "language games" in which the speakers were engaged) is essential in order to reach meaningful and reliable conclusions about the data.

Language use in Dialog Contexts

When people communicate, they ordinarily construct and ground joint perspectives (Clark & Wilkes-Gibbs, 1986). Upon re-referring to something, they typically mark that they are talking about the same thing by re-using the same (or a shortened) referring expression; this is known as *entrainment* (Brennan & Clark, 1996; Garrod & Anderson, 1986). Entrainment has been observed in many laboratory studies by psychologists (e.g, Gergle, Kraut, & Fussell, 2013; Horton & Gerrig, 2005; Isaacs & Clark, 1987) as well as in analyses of speech corpora by computational linguists (e.g., Nenkova, Gravano, & Hirschberg, 2008's studies of

the *Switchboard* and *Columbia Games* corpora, and Stoyanchev & Stent, 2009's from *Let's Go*, a working public information system about bus schedules).

A paper describing the corpus, methods, and tasks in greater detail (Brennan, Schuhmann, & Batres, 2013) is presented as a poster and is available in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 2013). There, we describe preliminary results from the laboratory tasks (Tasks #2-5 described in upcoming Method section). Those results included significant rates of lexical entrainment across the 6 rounds of the matching task, evidence for partner-specificity in referring expressions, and strong influence on navigation efficiency by direction-givers' (as opposed to direction-followers') spatial ability.

For the current workshop, we extend these findings (ibid) with newly coded data from the navigation task (Task #1, the Walking Around Corpus) in order to detect the effects of collaborative navigation on lexical entrainment and variability in referring expressions. For our purposes, the corpus provides a testbed for lexical entrainment "in the wild" and, in conjunction with the laboratory tasks, provides a way to understand how perspectives and components of referring expressions persist or evolve and change across contexts. We expected to find strong influence of previous contexts on subsequent ones. However, an *audience design* hypothesis predicts that even though prior experience shapes behavior and processing, contextual constraints can be applied flexibly, with speakers and addressees adapting "for" their partners when partners' needs are understood.

Basic and Applied Objectives

We aimed to collect a corpus that met these objectives:

1. To systematically capture the substantial variation in spontaneous speech and language use in a real-world environment, using several dozen pairs of speakers.
2. To better understand some key ways in which speakers tailor utterances to particular addressees in real time.
3. To parameterize the corpus for testing hypotheses about effects of visual context on spatial expressions.
4. To enable corpus data to be correlated and aligned with other data (e.g., time elapsed, pedestrians' locations, speakers' individual differences, performance data).
5. To enable qualitatively describing and quantitatively modeling convergence on shared vocabulary in discourse, or *entrainment*, across several different kinds of contexts (distinct environment for each remote partner, collaborative navigation in a real-world context, solitary recall, the effect of encountering a new speaker who uses different terms, multiple rounds of a laboratory-based referential communication task).
6. To provide useful information for modeling variation in human language use, with the goal of improving automated spoken dialog systems.
7. To enable characterizing how GPS navigation for pedestrians should differ from that for drivers, and to capture a range of strategies people spontaneously use in giving and following pedestrian directions.
8. To ensure that the corpus is of sufficiently high quality, in both audio recording and transcription, that it will be useful to other researchers.

Our current focus is to examine how perspectives and the referring expressions that index them evolve and change across contexts. This is relevant to the question of audience design, in that language processing may differ in the presence of one partner vs. another partner and in a solitary context vs. in an interactive social context.

Method

Procedure

The sequence of tasks that each pair completed over a 2-hour period is schematized in Figure 1 and can be summarized as follows:

#1 **Navigation task:** The Giver (G) described a sequence of 18 locations (depicted in a booklet and on a map) to the Follower (F), who interacted freely with G, visited each location, and took a photograph of it.

#2 **Spatial ability test**: After F returned to the lab, each partner separately completed a timed mental rotation test (Card Rotation Test–S-1 [Revised] by Educational Testing Services, 1962, 1975).

#3 **Memory test:** Each partner separately viewed pictures of 36 locations (the 18 from the navigation task and 18 new locations) and responded "old" or "new" to each; they then recalled a name for each item judged "old".

#4 **Interference task:** The Experimenter used potentially competing labels while having F identify the photos taken while walking around.

#5 **Matching (referential communication) task:** Each pair matched identical copies of the set of photos taken by the pedestrian partner. F acted as director for Rounds 1-3 with G as matcher, and G acted as director for Rounds 4-6 with F as matcher.

#6 **Questionnaire:** Each subject individually completed a final questionnaire to self report such information as their confidence with maps and directions, their familiarity with campus, whether they drove, and whether they were acquainted with their partner. Few pairs were previously acquainted.

Private contexts

Each member of a pair brought systematically different influences to the task, as follows: G, the stationary partner, sat alone in the lab (but could ask the nearby Experimenter for help with any technical difficulties). G's computer screen displayed a marker on a campus map showing the locations for F to visit (in order), along with a GPS cursor showing F's approximate location. G also had a binder with a page for each location, each with an initial label and either a single close-up picture of the location, or both the close-up and a wider angle picture that provided more visual context (see Figure 2). Type of visual context alternated for each location and was counterbalanced to 2 lists. F, the mobile partner, was immersed in the natural outdoor

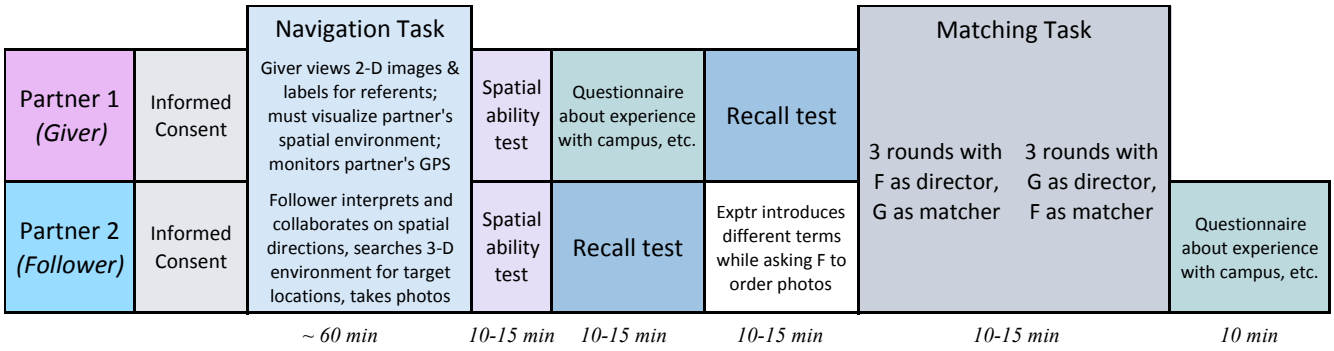

Figure 1: Timeline of tasks for each pair of partners; each session took approximately 2 hours.

environment; F photographed each location, and these images were used later for the matching task back at the lab.

Patriotic faces (aka profiles plaque)

Warning sticker (aka water tower sticker)

Mushroom house (aka brown brick structure)

Figure 2: Visual context: For each location, G saw a close-up view (left), with or without a wider view (right)

F's private context included an interference task (after the navigation task and prior to the matching task) in which the Experimenter introduced potentially competing labels (that differed from those given to G) while asking F to identify printouts of the photographs taken while walking around. This task, and the solitary memory and spatial tests, represented distinct intervening contexts that provided an opportunity to examine the extent to which partners would achieve shared perspectives or *conceptual pacts* (Brennan & Clark, 1996) during the matching task that incorporated the perspectives they had discussed earlier while navigating.

Transcription and Coding

Walking Around Corpus

We transcribed 36 dialogs from sessions in which both partners had signed an additional release (beyond the informed consent form) that allow us to make their data available to the research community. The digital recordings were transcribed in detail (including speech disfluencies, fillers, estimates of silences, and time stamps) and checked for accuracy and consistency.

For the purposes of the current study, the corpus was then segmented into stretches of speech associated with directing F to walk to each of the 18 target locations. We highlighted material in the transcripts (predominately nouns and adjectives) that givers and followers had used to refer to or describe target locations. This we used to determine how the labels provided to givers in the binders may have shaped the referring expressions they used to propose perspectives to followers during the spontaneous navigation task, as well as to identify the sources of material in referring expressions produced in both the (solitary) recall test and the (collaborative) matching task later on.

Memory Test

We recorded the time it took for G and F to respond that a close-up picture of a location was either *old* or *new*. The expressions they recalled for old targets were transcribed and coded for the sources of the words in each expression.

Matching Sessions: Entrainment Coding

For the findings we report in the CogSci '13 Proceedings (Brennan et al., 2013), we transcribed and coded the 6 rounds of the matching task for lexical entrainment (whether partners converged on the same referring expressions for a target location in Round 6, after having referred to the same location in Rounds 1-5). By our strict criteria (adapted from Brennan & Clark, 1996; Bortfeld & Brennan, 1997), all of the content words (nouns, adjectives) in the expression produced by the director (G) in Round 6 had to have been

included in at least one of Rounds 1-3 (where F was director), *or* had to have been explicitly confirmed (e.g., by a repeating them or by verbal acknowledgment) by F (the matcher) in Rounds 4-5. In other words, there needed to be evidence that both partners ratified all elements in the expressions in Round 6 (so "ship sculpture" was not coded as convergent with "ship structure", nor "faces sign" with "faces plaque"; see Brennan et al. 3013 for detail and examples). However, entrained-upon expressions in Round 6 were often shortened from earlier expressions.

Our coding to date has focused on propositional content, ignoring determiners, prepositions, quantifiers, and the order of content words, as well as proxy terms (e.g., "*thingy*", "*blah blah*"), and hedges and disfluencies (e.g., as when G proposed: "*it's like a (.) a brick (.) statue square-looking thing with like a mushroom cut-out in the middle*" (Pair 14). However, these elements are interesting in their own right, and may be related to strategies of spatial direction giving and possibly to individuals' spatial ability. Future coding of referring expressions will explore such possibilities.

In order to test our *audience design* hypothesis (that speakers design utterances "for" specific addressees), we coded those 10 target destinations for which the Experimenter had provided a distinct competing expression to F before the matching task (see white box in Figure 1, above). For each of the 6 rounds of the matching task, we coded the **source context** for all the elements of each target referring expression across all contexts of the session:

- *Given*: Material provided originally to G in binder
- *WAC*: Material produced by either G or F during navigation task (highlighting *Given* material)
- *Exptr*: Potentially competing elements produced by the Experimenter. Some of these happened to be introduced during the navigation task (in which case *WAC* was considered the source).
- *Other*: Elements that were not produced in any referring expression prior to the matching task.

A referring expression could contain elements from any combination of these sources. An expression from the recall task could contain any elements except *Exptr*, the influence of which occurred after the recall test.

Results and Discussion

Here, we incorporate some new data from the corpus of navigational dialogs from the Walking Around Corpus with the results from the 49 pairs of subjects who did the laboratory tasks we presented earlier (Brennan et al., *2013)*. Coding is ongoing (with 1/4 of WAC coded in detail), so we report descriptive statistics without hypotheses tests.

Influence of G's and F's Distinct Contexts

The navigation task differed from the usual referential communication task in several important ways. When first encoding the target locations, G and F inhabited distinct perceptual contexts (even though they did the navigation task together). G was in the lab prompted by a binder with 2-dimensional pictures and a label for each location; F was immersed in the 3-dimensional world, moving through space. G could monitor F's progress via GPS on a map of the campus, updated every 2 seconds; however, unlike in a driving task, G could not assume that F was facing in any particular direction, or that F was traveling along a predictable path. Moreover, walking is slower and pedestrian landmarks are on quite a different scale than those useful in driving. To accomplish the task efficiently, F had to take responsibility for communicating her focus of visual attention as well as local landmarks to G. G, on the other hand, put much effort into visualizing what F could see, based on F's location on the map. Both partners were to some degree familiar with the campus, although not uniformly so (and targets consisted of relatively unfamiliar details and odd corners of campus). The dialogs show much checking of one another's knowledge (more so than the typical lab-based referential communication task).

Initial Strategies G, with plenty of time to encode and reflect on target labels in the stimulus binder while waiting for F to walk around, began by proposing this expression to F 53% of the time. This was especially true (58% of the time) when G had only a close-up view of the target. Often, however, this proposal was not taken up by F, and they were just as likely to end up entraining on a different perspective by Round 6 of the matching trials (see 1).

We expected that G would find it easier to effectively propose a location description to F when G had more visual context. Because most targets were insignificant details on campus, many of the close-up images were unfamiliar and hard to describe without a wider-angle visual context. We are currently coding start and end times for each target to detect any effects of visual context.

Table 1: Potential effects of G's 1st proposal to F during navigation task upon later expressions *(N=9 pairs)*

Likelihood of:	G proposes verbatim label from binder	G proposes novel or tentative expression
G's later correct recognition	.91	.86
F's later correct recognition	.79	.86
Entraining during matching task	.60	. 63
Entraining on given label (verbatim) during matching	. 24	.30

Spatial Ability (Task #2)

Given the differences between partners' task roles and private contexts, it makes sense that spatial ability mattered more in the giver role than in the follower role. For the 54 pairs, the time it took to complete the navigation task was reliably correlated with G's mental rotation ability, $r = -.42$, $p < .005$, but less so with F's mental rotation ability, $r = -.24$, $p = .11$.

Memory Test (Task #3)

Performance Because memory prompts consisted of the same close-up photos from the binder initially provided to G, the context in which G encoded the destinations was consistent with the context of recall. This was far from the case for F, who had to encode the locations by discovering them out in the world, often from a very different vantage point than from G's binder photos (this was evident from the photographs F took). In the small sample of 9 pairs, G successfully distinguished "old" from "new" targets 88% of the time, while F did so 70.8% of the time. In the larger sample of subjects (Brennan et al., 2013), G did better than F (with 90.6% correct vs. F's 80.4%, $t(47) = 3.69$, $p = .001$). In that sample, G also responded marginally faster than F, $t(47) = 1.72$, $p < .10$.

Convergence of G's and F's Recalled Terms After responding "yes" in the recognition test, G produced an expression with the same content as F 41.7% of the time (N=49 pairs). The fact that this was so much lower than their entrainment rate in the matching rounds later on highlights the difference between a context of individual recall and one in which perspectives are jointly-achieved.

Effects of Interference (Task #4) on Audience Design in the Matching Task (Task #5)

We coded the extent to which competing expressions that were introduced by the Experimenter to F in Task #4 (and that were *not* mentioned by G or F in the dialogs while walking around) affected the expressions used later during the matching task (#5). This is graphed in Figure 3 (dark bars), along with the influence of elements from the labels originally provided to G in the binder (light bars) and the influence of elements mentioned while walking around (medium bars). Recall that G and F's navigational dialogs often contained elements from G's given labels. Note that the greatest influence on referring in the matching rounds comes from G and F's prior interaction while walking around, with relatively little from the Experimenter (which was the most recent expression F heard just before matching pictures with G). In fact, the referring expressions F addressed to G in Round 1 included material from the Experimenter's expressions (that F had never before used with G) only 25% of the time, compared to elements in common ground from the navigation task over 80% of the time (note that these do not add up to 100% because an expression could be composed of elements from multiple sources). This is strong evidence that F kept good (if not perfect) track of the perspectives established with G.

Note that although the influence of the interfering (Experimenter) partner diminished somewhat during the matching rounds, G accepted many of F's proposals in Rounds 1-3 and started using the Experimenter's terms while acting as director in Rounds 4-6. This is evidence that conceptual pacts are flexible.

The discontinuities between the bars for R1-3 and R4-5 in Figure 3 were due primarily to F and G switching director-

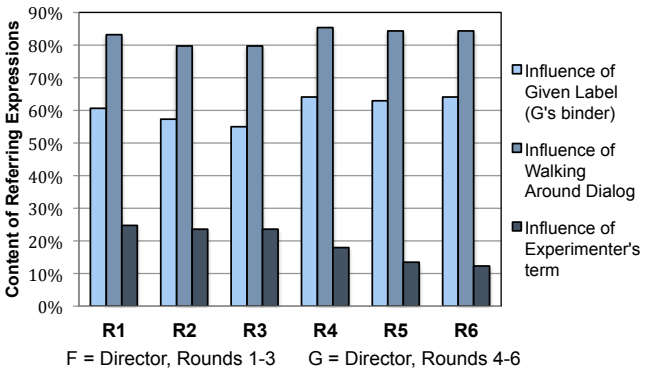

Figure 3: Three sources of influence on expressions used in matching task: Labels orig. provided to G, dialogs while walking around, and Exptr's competing terms (N = 9 pairs)

matcher roles; to the extent that a perspective was modified during the matching rounds, it was most likely to happen then. Many of the failures to entrain by Round 6 were due to G producing (in a longer expression) one element that F had not used in Rounds 1-3 (or confirmed in Rounds 4-6). Note also that G (as director in R4-6) shows a slightly greater influence from the originally given labels and from the navigation task (recall that those sources are not entirely independent) than F does (as director in R1-3).

Shared contexts differ from solitary contexts Often, the term F proposed in R1 was not the same as the one F recalled in Task #3; likewise, the perspectives G and F entrained on in Task #5 were often not the same as G's and F's recalled terms in Task #3. This suggests that speakers differentiated between shared and solitary contexts of referring. Moreover, the expressions that pairs entrained on in the matching task were frequently based on perspectives proposed during the Walking Around task (even though these tasks were separated by (solitary) spatial ability and memory tests). This is particularly striking when perspectives were unusual, as in these examples of two different pairs discussing the same location. In both cases, G began with the binder label; overlaps are marked with * and pauses are marked with (.):

G: you're looking to take a picture of like (.) you know those warning sticker things they have on the the tanks?
F: okay
G: you're taking a picture of that and it should have like a blue three a red zero a yellow zero and a white *S A*
F: okay
G: so basically Uno cards *{both laugh}*
G: *I don't know why they remind me* of Uno cards
F: *okay that's* (.) that's totally a good way to think of it *{more laughter}* *(Pair 3, # 6)*

Later while matching in R1, F proposed "*the weird, UNO-looking sign, 3 0 0 3*" and shortened this to "*UNO cards"* in R2-3; G further shortened this to "*UNO*" in R4-6.

In another pair, G began similarly with "warning sticker" and a lengthy description, to which the partner responded:

F: looks like a ha- (..) ha-, looks like one of those hazmat stickers {*unintelligible*} *right?*
G: *yes exactly* yes it's like a hazmat sticker *(Pair 19, #6)*

Later, these two entrained on "*hazmat*" while matching. Note that during the memory tests for both of these pairs (3 and 19), neither G nor F produced the terms from these conceptual pacts, and both givers produced the label "warning sticker" from the binder. Detailed coding continues and will determine whether this pattern holds true for the rest of the corpus.

Conclusions

Preliminary analyses of the language used in these multiple, multi-modal contexts demonstrates that conceptual entrainment occurs even when members of a pair initially encounter targets without having similar visual co-presence (so targets had quite different figure/ground relationships). Givers saw colored 2D photographs of labeled targets and locations on a map, along with additional spatial context for half of them, while followers learned about the targets from their partners and then had to discover them within a 3-dimensional real world environment (with many visual viewpoints possible). Our subjects used language in a way that suggests they kept track of partner-specific context (addressing relatively few of the Experimenter's terms to their partner, even though they heard these more recently), and that they distinguished joint contexts from solitary ones.

At the same time, these data highlight the flexible nature of conceptual pacts. We saw ample evidence that speakers made proposals to their partners about how to conceptualize a target, marking their commitment to these proposals (e.g., with hedges), along with evidence of partners ratifying or modifying such proposals or making counter-proposals. When partners reunited in the second joint task and when one introduced new material, it was often (if not always) the case that the other quickly adapted.

These findings suggest that speakers are able to maintain and keep track of multiple perspectives about the same items that are linked to different contexts or task goals. Memory does not oblige them to reflexively produce terms that have been recently primed; apparently, context (including an interacting partner) provides cues that allow them to adapt in partner-specific ways. This holds implications for the psycholinguistics of dialog and for computational models of the representation and processing of spontaneous referring expressions in spoken dialog.

We continue to code referring expressions in this project, with the goal of providing empirically grounded conditional probabilities that can relate expressions used by the same speakers across different contexts. To the extent that computational models seek to characterize variability, to provide a foundation for spoken dialogue interfaces that limit or shape such variability, and to predict what forms of expressions or disfluencies may be likely in upcoming utterances, the Walking Around Corpus and the associated database may be a useful resource. Please contact susan.brennan@stonybrook.edu for more information.

Acknowledgments

This research was supported by the National Science Foundation under Grant #IIS-1043665. We thank M. Walker and Z. Hu for helpful discussions, R. Bowmani for technical assistance, and S. Eng, L. Snelling, R. A. Wang, and M. Zhang for assistance with coding and transcription.

References

Bortfeld, H. & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and nonnative speakers. *Discourse Processes, 23*, 119-147.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 6*, 1482-1493.

Brennan, S. E., Schuhmann, K., & Batres, K. (To appear, 2013). Entrainment on the move and in the lab: The Walking Around Corpus. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.

Galati, A., & Brennan, S. E. (2013). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language and Cognitive Processes*.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 181-218.

Gergle, D., Kraut, R.E., & Fussell, S.R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction, 28*, 1-39.

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1-35.

Hwang, J., Brennan, S. E., & Huffman, M. K. (submitted). Disambiguating phonetic ambiguities in non-native spoken dialogue: Effects of priming and audience design.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*, 26–37.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proc. of ACL/HLT* (Short Papers), 169-172.

Stoyanchev, S. & Stent. A. (2009). Lexical and syntactic adaptation and their impact in deployed spoken dialog systems. In *Proc., NAACL HLT* (Short Papers), 189-192.