What Tunes Accessibility of Referring Expressions in Task-Related Dialogue?

Ellen Gurman Bard (ellen@ling.ed.ac.uk)
Linguistics and English Language, University of Edinburgh, Edinburgh EH8 9LL, UK

Robin Hill (r.l.hill@ed.ac.uk)
Human Communication Research Centre, University of Edinburgh, Edinburgh EH8 9LW, UK

Mary Ellen Foster (foster@in.tum.de)
Informatik VI: Robots and Embedded Systems, Technische Universität, München, Germany

Abstract

Ariel (1988) proposes that the grammatical form of any referring expression can be predicted from the deemed accessibility of its referent to the intended audience. The term 'deemed' is critical: it allows the speaker an egocentric perspective and frees choice of expression from the actual contingencies of the situation in which it is uttered. We analyze 1775 first mentions of visible objects within a multi-modal corpus of cooperative task-related dialogues (Carletta et al., under revision) for effects of situation (communication modalities, actions involving the named entity) and of responsibilities assigned to speakers and listeners. Accessibility distributions show statistically significant effects of three kinds: circumstances readily available to the listener (concurrent movement of the named object); circumstances private to the speaker (hovering the mouse over the object, when the listener cannot see the mouse), and speakers' assigned roles.

Keywords: reference, accessibility, corpus experimental studies, pragmatics, situated dialogue

Introduction

The question of what a thing will be called engages everyone interested in the interpretation and generation of referring expressions. One very wide-ranging approach, (Ariel, 1988, 1990, 2001) attempts to key elaboration of the form of referring expressions to the 'deemed' a priori accessibility of the referent, that is, to how difficult the producer of the expression estimates it will be to access the referent concept, discourse entity, or extra-linguistic object. Expressions introducing entities deemed completely unfamiliar to the audience should be maximally detailed, as in, for example indefinite NPs including modifiers of various kinds ('a former Republican governor of strongly Democratic Massachusetts'). Expressions of intermediate accessibility might be definite NPs, deictic expressions, or personal pronouns in that order. Expressions making reference to a unique, just mentioned entity in focus can be as minimal as so-called clitics, unstressed and all but deleted pronouns ('/z/ in the garage'). Accessibility theory provides a unified framework for predicting how forms of referring expressions will respond to givenness, discourse focus, inferrability from local scenarios and the like. As a general notion, accessibility ought to include effects of any available conditions which might draw attention to the correct referent. This paper discusses two such conditions, task related movements and the roles of the players in a collaborative task.

Our questions about these conditions hinge on the information which human interlocutors might use in determining how to refer. Ariel's notion of accessibility appears to depend on what the speaker supposes is the case, not on what is genuinely easier for the listener. While some approaches to dialogue assume that speakers carefully model their interlocutors, so that initial forms of expression could arise from the interlocutors' needs (Brennan & Clark, 1996; Clark & Krych, 2004; Schober, 1993), there is increasing evidence that we have limited ability to construct, recall, or deploy any such model in a timely fashion (Bard et al., 2007; Horton & Gerrig, 2002, 2005a, 2005b; Horton & Keysar, 1996). Interlocutors may behave egocentrically (Bard et al., 2000; Bard & Aylett, 2004; Horton & Keysar, 1996), adopt a global account of affordances of a situation, (Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon, 1997; Brennan, Chen, Dickinson, Neider, & Zelinsky, In press), or observe information indicative of the listener's knowledge, but fail to act on it (Bard et al., 2007; Brennan et al., In press).

The situation for form of referring expressions is mixed. While accessibility of referring expressions is more sensitive to the knowledge of the listener than is clarity of articulation (Bard & Aylett, 2004), other studies show that tendencies to match nomenclature to listener's history or current situation are quite variable (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005a; Horton & Keysar, 1996; Keysar, Lin, & Barr, 2003). So-called conceptual pacts are actually lexical pacts (Brennan & Clark, 1996), agreements to call objects by certain names, and are the result of negotiation over time, across which accessibility of the referring expression naturally rises. If speakers do track one another's internal states, the accessibility of even introductory mentions will suit the interlocutor's current needs, rather than the speaker's.

The evidence may be inconclusive because the typical paradigms for dialogue studies restrict cooperation to disjointed episodes. In typical tasks, one participant instructs another to act on or select from an array of potential referents, while the other follows instructions relative to an identical or partially overlapping set. Both responsibilities and activities are clearly distinct; the channels for

communication are purposely limited; and the knowledge shared between instructor and listener is altered trial by trial in an unpredictable way. To discover whether more robustly cooperative behaviour appears in more cooperative tasks, we have created a corpus of dialogues centered around a shared task which demands joint attention and joint planning.

The JAST project, which studies joint action in the hopes of finding human models for cooperation in robots, has developed a Joint Construction Task, in which two human players cooperate to construct a tangram in a shared workspace represented on their yoked screens (Figure 1a). Each player can manipulate the component parts by mouse actions. Each dyad works under one of two role conditions: either one player is assigned to manage the task and the other to assist or no roles are assigned. Mouse actions draw attention both because they are integral to the construction process, and because, to mimic the industrial risks of working with robot partners, they are dangerous. If both mice grasp the same object, it breaks and must be replaced. If two objects overlap, both break. Because players act on the tangram parts and sub-constructions, the activity of grasping or moving the named object adds a haptic or praxic modality to spoken forms. Even 'hovering' the mouse over a part without grasping it offers a chance to make a part accessible.

To discover how well keyed any change in form of referring expression is to the perceptions of the interlocutor, the design contrasts situations in each players' mouse cursor is cross-projected onto the other's version of the shared screen and situations in which each player see only the movement of the object which the other 'grasps'. Only in the first case can a player see the mouse 'hovering' over a tangram part which is not actually moving.

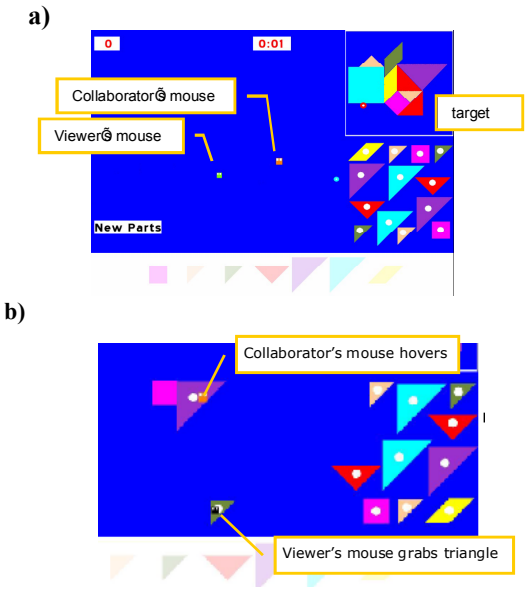

Figure 1 Joint Construction Task shared screen a) at outset; b) during construction (detail)

If moving a part draws attention, it should also draw referring expressions of greater accessibility. Since pointing is associated with shorter, less detailed referring expressions and pointing to closer targets has an even stronger effect (Kranstedt, Lücking, Pfeiffer, Rieser, & Wachsmuth, 2006), touching and moving should have a very marked effect on the form of expression. Like Kranstedt et al., we note the association of the 'hand' location and verbal deixis: in our case a larger proportion of verbal deictics (*this square; these, mine*) than of other forms coincide with mouse-referent overlap (Foster et al., 2008). To go further, we will divide overlaps by whether players actually move parts.

If the listener's knowledge is of concern, moving parts should have different effects on accessibility from merely leaving the mouse hovering over them. Since movements of objects will always be visible to the listener in the Joint Construction Task, a speaker adjusting to listener knowledge could use deictic forms to refer to currently moving parts. In contrast, visibility of the mouse cursor should determine whether hovering is shared information: referring expressions with a visibly hovering mouse should be made more accessible in form than those with an invisibly hovering mouse. In fact, a speaker might even increase the accessibility of an expression referring to a part which the *listener* is touching or moving. Again, responses should depend on what is visible. If the listener's knowledge is less important than the speaker's, however, the speaker's own hovering movements should attract higher accessibility forms regardless of what the listener can see.

The players' roles suggest further questions. Managers have a primary role in setting the dyad's agenda. They should have more power to designate discourse focus and to change it, for example. If the choice of a level of accessibility is an overt designation on the speaker's part, then managers should have special powers of designation. And, as we suggested earlier, managers might have less reason to track or adjust to the needs of their partners than role-less or assistant players do. Alternatively, assistants may feel obligated to conform to patterns established by managers more that equal partners would.

In all cases the answers to our questions should be reflected in distributions of referring expressions across ordered levels of accessibility. Though accessibility bears on the relationships between earlier and later mentions of an entity, it ought to be important to determining the form of introductory mentions, too. By restricting our investigation in this way, and by controlling the objects available for naming, we can test our hypotheses about how a thing shall first be called.

Corpus Collection And Coding

Task

The Joint Construction Task or JCT (Carletta et al., under revision) offers to two collaborating players a model tangram (see Figure 1a, top right), geometrical shapes for reproducing it (center right), a work area (center screen), a counter for breakages (top left), a set of replacement parts (bottom of the screen), and a clock measuring elapsed time (top center). The players' task is always to construct a replica of the model tangram as quickly, as accurately, and as cheaply in terms of breakages as possible. An accuracy score is provided at the end of each trial.

Participants manipulate objects by left-clicking and dragging them or by right-clicking and rotating them. Carefully timed collaboration is required. Any part or partially constructed tangram 'held' by both players will break and must be replaced from the spare parts store to complete the trial. Objects can be joined only if each is held by a different player. Moving an object across another breaks both. Objects join permanently wherever they first meet. Inadequate constructions can be purposely broken and rebuilt from spare parts, incurring a cost in parts and time.

Figure 1 shows the shared portrayal of the game state but includes objects that may be visible to only one player. At start of play (Figure 1a), the viewer's and the collaborator's current mouse positions are represented by an orange cursor and a green cursor respectively. A small blue circle marks the collaborator's current gaze position.

In a magnified view of a later point in a different trial, Figure 1b shows how grasping an object is distinguishable from mere superimposition (hovering). The viewer's mouse cursor retains its original colour while hovering over a partially constructed tangram, but the listener's cursor is shown in black when it has grabbed a green triangle.

Apparatus

Each participant sat approximately 40cm from a separate CRT display in the same sound-attenuated room. Participants faced each other, but direct eye contact was blocked by the two projection computers between them. Participants were eye-tracked monocularly via two SR-Research EyeLink II head-mounted eye-trackers, but Eyetracking results are not relevant to the current report. Headworn microphones captured speech on individual channels. Continuous audio and video records were kept, including a full record of locations and movements of individual parts, constructed objects, and cursors. Composite Camtasia videos recorded all movements and audio.

Participants, design and materials

Sixty-four Edinburgh University students, paid to participate, were paired into 32 same-sex dyads who had never met before. Four further dyads were discarded because of technical failures. Each dyad participated in 8 experimental conditions produced by the factorial manipulation of three communication modalities: speech, gaze (each player's current eye-track cross-projected onto the other's screen twice within each 42 ms cycle), and mouse cursor (also cross-projected). Participants could always see their own mouse cursor. Without no additional communicative modalities they saw only the moving parts. Gaze and mouse conditions were pseudo-randomised following a latin square. Speech was allowed in the first four presented conditions for half the dyads and in the second four for the rest. Only conditions with speech are analyzed here.

In 16 dyads one participant was randomly designated Manager and the other Assistant. The manager was asked to maintain "Quality Control", in speed, accuracy, and cost, and to signal the completion of each trial. The assistant was to help. The remaining dyads were assigned no roles but otherwise had the same working instructions. Trials ended when one player declared the construction complete by pressing the keyboard spacebar and the other confirmed. An accuracy score reflecting similarity between the built and the target tangrams then appeared across the built exemplar.

Each dyad reproduced 16 different target tangrams, 2 for each condition. No tangram resembled a nameable object. Each contained 11 parts. At the beginning of each trial, the same set of 13 parts was available, comprising 2 copies of each of 6 shape-color combinations (squares or right-angle isosceles triangles differing in size and colour) and a single yellow parallelogram. The 13 parts appeared in 4 different layouts counterbalanced across experimental items. The extra pieces differed from trial to trial.

Table 1 Accessibility Coding Scheme

Level	Definition	Examples
0	Indefinite NP	*I'm grabbing a purple one* *one of the nearest blue pieces*
0.5	Bare nominal	*Pink one ?* *Triangles*
1	Definite NP	*I'll put the red bit on.* *The other purple one.*
2	Deictic NP	*Okay now just those two little kids.*
2.5	Deictic Poss } Pro	*Do you see where mine's moving now?* *Now we need to fit these.*
3	Other pronouns	*Can you hold it?* *You wanna just bring it up?*
4	Clitic/inaudible	*/s/ not going to fit.*

Coding referring expressions

Dialogues were transcribed orthographically. Each referring expression was time-stamped for start and end points. Then, each expression referring to any on-screen object was coded

with a referent identifier linking it to the object. Coders had access to the video and audio track and were allowed to use any material within a dialogue to determine the referent of any expression. All referring expressions were coded for accessibility on the scale given in Table 1. This system represents a modest expansion of a system applied to an earlier corpus of task-related dialogues (Bard & Aylett, 2004) and yielding negligible disagreement between coders.

Results

Overall outcomes

Figure 2 presents the overall distribution of first mentions across the accessibility scale. Despite the fact that all but one of the original parts had an identical competitor and might invite an indefinite referring expression ('a small red triangle'), only 16% of first mentions were indefinite NPs. The remaining 84% were of higher accessibility.

Very accurate tangrams were built in all conditions. The cross-projection of mouse location was helpful. Trials with cross-projection were significantly shorter on average (187sec v 205sec: F = 11.45, df = 1,30, p = .002) and incurred significantly fewer breakages (1.8 v 2.3; F = 4.52, df = 1,30, p = .008) while achieving equal accuracy (92.1 v 91.9). Managerial structure was neutral in ultimate outcome. While Manager-Assistant trials were significantly longer than No Role trials (216sec v 175sec; F = 10.67, df = 1,30, p = .003), they gave similar performance (Accuracy: 93.8 v 91.2; Breakages: 2.0 v 2.1).

Modalities, knowledge, and accessibility

Method. The conditions critical to our predictions were coded for a multinomial logistic regression which modelled the distribution of first mentions across accessibility categories. This statistic can test the capacity of category variables (like Mouse v No Mouse) to influence ordinal variables (like accessibility). It does so by constructing regression equations both for the whole ordinal series and for the comparison of each level in the series to some reference level. We will use it to ask which of the critical differences in action and communication change the tendency to produce indefinite referring expressions relative to the each other category.

The calculations are describe odds ratios, but for interpretability, we display simple proportions of cases. To eliminate empty cells, accessibility categories were collapsed into four levels: Indefinite NPs (including bare nominals), Definite NPs, Deictics (including deictic NPs, deictic pronouns and possessive pronouns) and Other Pronouns (including clitics). Uninterpretable or disfluent items (less than 1% of the data) were omitted.

Separate equations were prepared for the Mouse Cursor Cross-Projected (n = 836) and No Mouse Cursor conditions (n = 939). The predictors included the experimental variable Roles Assigned, the participants' mouse actions (the speaker/listener moving part being mentioned, or 'hovering' the mouse over it), and the interactions of Roles Assigned with each of the movement variables. Gaze cross-projection was not included, as it had proved an ineffective predictor in earlier exploratory regressions. Table 2 shows the significant outcomes. Each effect listed is essentially independent of any effect from any concurrent predictor variable.

No effect of the listener's actions reached significance. Instead there were effects of the speaker's actions and of Role Assignment As predicted (Figure 3), visibly moving the referent increased deictic expressions (31 v 46%) at the expense of indefinites (18 v 12%) whether or not the speaker's mouse was visible. Other mouse movements also affected accessibility: A visibly hovering mouse cursor (Figure 4) accompanied a significant shift to deictics (39 v 51%) from pronouns (15 v 6%), while an invisibly hovering mouse decreased both indefinites and pronouns in favor of deictics. Role assignment influenced the effects of invisible mouse gestures: Only in Manager-Assistant dialogues, did introductory mentions shift markedly away from indefinites (22 v 8%) toward deictics (25 v 40%) with hovering, much as they did with visible mouse gestures. In No-role dialogues definites predominate in either case.

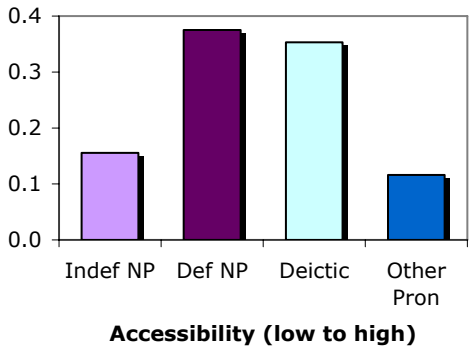

Figure 2. Accessibility of first mentions

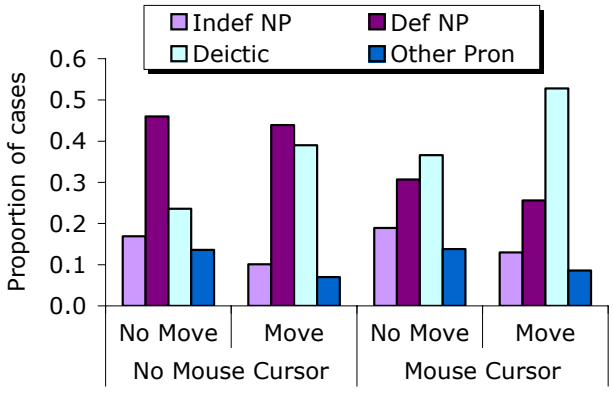

Figure 3. Accessibility of first mentions: Effects of speaker moving referent object

Table 2 Significant predictors of accessibility. For individual levels of accessibility, *df* = 1. * = *p* < .05; ‡ = *p* < .01; § = *p* < .001

No Mouse Cursor Cross-Projection						
	-2 Log Likelihood		χ^2	*df*	Cox & Snell	
	268.07		105.00§	27	0.106	
	Speaker Move		Speaker Hover		Speaker Hover x Roles Assigned	
χ2					276.00	7.42*
	B	Wald	B	Wald	B	Wald
Definites			-1.137	10.85§	1.075	4.99*
Deictics	-0.814	4.78*			1.275	6.17*

Mouse Cursor Cross-Projected						
	-2 Log Likelihood		χ^2	*df*	Cox & Snell	
	258.00		61.34§	27	0.071	
	Speaker Move		Speaker Hover		Speaker Hover x Roles Assigned	
χ2					266.00	7.77*
	B	Wald	B	Wald	B	Wald
Deictics	-0.722	5.05*				
Pronouns			1.264	6.95‡		

Discussion

This paper asked whether the association between handling a thing and using an accessible format to name it was linked to the speaker's own knowledge or to the knowledge expected to reside with the listener. There are two reasons to believe that the listener is not in charge. First, we found no significant effects of the listener's manipulation of tangram parts on the speaker's form of referring expression, even when the listener's movements were fully visible to the speaker. Second, we did find effects of speakers' mouse gestures which invisible to the listener.

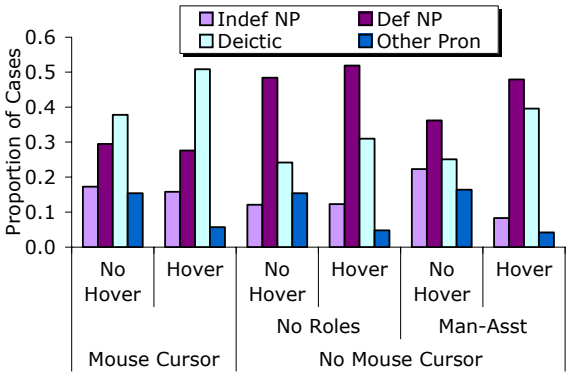

Figure 4. Effects on accessibility of hovering mouse over referent, by mouse cursor visibility and assigned roles.

At the same time, we suggested that if accessibility is an expression of opinion, it should be manipulated by managers in particular. In the event, Manager-Assistant dialogues showed more egocentric use of private gestures than No-role dialogues. Such gestures accompanied a shift to deictic expressions, as they did when mouse cursor projection made them visible.

Though the results support the general predictions, they invite further interpretation. We see three major issues.

First, the results fall some way short of a clear case for managerial insensitivity. The Role Assignment results were based on expressions produced by both participants. Analyses comparing managers with assistants are made difficult by small or empty cells. Both show that the pattern shown in Figure 4. Accordingly, we have no particular evidence contrasting managers with their assistants, though we can distinguish them from the dyads who had no assigned roles.

As we suggested earlier, however, one result of social inequalities is to give precedence to one individual. The manager decided what should happen next. To cooperate, the assistant had to conform to the manager's choices. Conforming to the manager's referential habits, for social reasons, or through structural priming, could make the assistant designate with invisible gestures, too. In essence, the assistant can achieve a tendency toward use of definites or deictics where they might otherwise appear to be unwarranted and then employ private gestures to accompany these instances. In contrast, role-less dyads might follow a mixture of styles or compete to control the task plan or the naming habits.

Second, though it is clear that speakers' private and public actions associate with particular levels of accessibility, it is not clear that their effects are all increases in accessibility. For example, Figures 3 and 4 show a tendency, significant only for hovering, for speaker actions not to collocate with the highest levels of accessibility in first mentions: pronominal or clitic introductory mentions are used less often when the mouse overlaps the referent part than when it does not.. Thus, the haptic or ostensive functions of mouse movements literally designate objects: they turn *a triangle* into *this* but they do not turn *this* into *it*. For this reason, the single accessibility continuum can be viewed as a set of referential phenomena, for example, demonstration or givenness in context, bearing on speakers' choices with different degrees of force.

Finally, there is the issue of the discourse history within which the introductory mentions are set. Clearly, some first mentions do not refer to totally discourse-new or completely unpredictable entities (Prince, 1981). There is no doubt that other forces work on the choice of referring expressions. Whether the choice of referring expression is always to some degree an indication of the speaker's view of accessibility we do not know. What seems likely in domains where public and privileged actions have similar effects is that it can certainly be a case of listener-insensitive designation at times.

Acknowledgments

This work was funded by EU Project JAST (FP6-003747-IP). The authors are grateful to the JCT programmers, Tim Taylor and Craig Nicol, Joe Eddy and Jonathan Kilgour for their contributions to the software, to Jean Carletta who managed the program development, to the reference coders for guessing what the speaker had in mind, and to JAST colleague J.P. de Ruiter for inspiring this exploration of the JCT corpus..

References

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics, 59*(4), 580-592.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics, 24*, 65-87.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London.: Routledge/Croom Helm.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text representation:Linguistic and psycholinguistic aspects.* (pp. 29-87). Amsterdam: John Benjamins.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language, 57*(4), 616-641.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Bard, E. G., & Aylett, M. P. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions.* (pp. 173-191). Cambridge, MA: MIT Press.

Brennan, S. E., Chen, X., Dickinson, C., Neider, M., & Zelinsky, G. (In press). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 1482-1493.

Carletta, J., Nicol, C., Taylor, T., Hill, R., de Ruiter, J. P., & Bard, E. G. (under revision). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods, Instruments, and Computers*.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–68.

Foster, M. E., Bard, E. G., Guhe, M., Hill, R., Oberlander, J., & Knoll, A. (2008). *The roles of haptic-ostensive referring expressioins in cooperative task-based human-robot dialogue.* Paper presented at Human Robot Interaction, Amsterdam.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47*(4), 589-606.

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

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127-142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H., & Wachsmuth, I. (2006). Deictic Object Reference in Task-oriented Dialogue. In G. Rickheit & I. Wachsmuth (Eds.), *Situated Communication,* (pp. 155-207). Berlin: Mouton de Gruyter.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-256). New York: Academic Press.

Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition, 47*(1), 1-24.