# Learning to Express Visual Contrasts in the Production of Referring Expressions in Yucatec Maya

**Lindsay K. Butler (lkbutler@email.arizona.edu)**
Department of Linguistics, Douglass Building, Room 200E
Tucson, Arizona 85721 USA

**T. Florian Jaeger (fjaeger@bcs.rochester.edu)**
Department of Brain and Cognitive Sciences, Meliora Hall, Box 270268
Rochester, NY 14627 USA

**Juergen Bohnemeyer (jb77@buffalo.edu)**
Department of Linguistics, 609 Baldy Hall
Buffalo, NY 14260 USA

**Katrina Furth (kfurth@bu.edu)**
Neuroscience, Boston University School of Medicine, 85 East Newton Street
Boston, MA 02118 USA

**Abstract**

We examined rates of informativeness in the production of modifications in response to a visual contrast In a video description task with speakers of Yucatec Maya. We analyzed modifications of referring expressions on the part of a speaker, and we also examined the effect of over- and under-informativeness on the listener's comprehension. We found that prior experience with difficult comprehension did not significantly affect the listener's rate of informativeness when in the role of speaker, but we found that experience as a speaker did result in reduced rates of under-informativeness. That is to say, as a speaker's own experience progressed, the speaker became less under-informative. We discuss these results as audience design-based learning.

**Keywords:** referring expressions, informativeness, audience design, learning, Yucatec Maya, field-based psycholinguistics

## Introduction

In order to successfully communicate their intentions, speakers need to select sufficiently informative referential expressions. The form that a referring expression can take ranges from a pronoun, like "it" to a lengthy descriptive noun phrase with modifiers, like "the donut with the pink frosting and blue sprinkles." Here we investigate to what extent speakers provide sufficient information to the intended reference to their audience and to what extent this depends on previous experience speakers had with the referential task.

Speaker-internal, language production processes clearly assert pressures on speakers and determine the form of the speakers' utterances. More accessible (animate, given or salient) referents have a privileged status in sentence production (Bock & Warren, 1985). They tend to occur sentence-intially (Bock, 1986; Ferreira & Dell, 2003; Ferreira & Yoshita, 2003) and they are more often assumed to be the topic of conversation (Hudson, Tanenhaus, & Dell, 1986; Hudson-D'Zmura, 1988; Gordon, Grosz, & Gilliom, 1993) and referred to in subsequent sentences with attenuated forms, such as pronouns (Prat-Sala & Branigan, 2000; Christianson & Ferreira, 2005). The reduction of accesible referents might be production-oriented or hearer-oriented.

There is evidence that speakers take into consideration the needs of addressees in designing their utterances. Speakers adapt their pronunciation based on the social identity of their addressee (Bell, 1984; Eckert, 2004) and make use of detailed assumptions about the common ground (Clark, 1996), engaging in what has been called audience design. Speakers have been found to use pronouns and attenuated forms when a referent is assumed to be in the focus of attention of the addressee (Brennan, 1995; Chafe, 1976, 1994; Gundel, Hedberg, & Zacharski, 1993; Prince, 1981). Speech rate has also been argued to involve some processing of addressee-oriented information (Galati & Brennan, 2010; Lindblom, 1990).

The extent to which speakers engage in audience design has been questioned. Speakers often fail to disambiguate utterances when they could do so by changing word order (Arnold, Wasow, Asudeh, & Alrenga, 2004) or by producing additional words (Ferreira, 2003; Ferreira & Dell, 2003). Galati and Brennan (2010) as well as Gregory, Healy, and Jurafsky (2001) do find sensitivity to addressee's knowledge, though adjust to speech rate is not perfect. Speakers have faster speech rates with novel addressee, for a story that the speaker has previously told. Bard et al. (2000) similarly find less clear articulation for the second mention of a referent, even if the first mention was not the speaker's own mention (for discussion see also Arnold (2008); Ferreira (2008); Bard and Aylett (2004)).

Another issue of debate revolves around when and how audience design is implemented. Speakers may initially produce sentences automatically ego-centrically then and only implement audience design when they have to (e.g. when given feedback from the listener, (Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998). Alternatively, speakers may go through a process of monitoring and adjustment (Horton & Keysar, 1996)

in which they begin with more ego-centrically designed utterances and adapt to the needs of an addressee over time. Recent evidence suggests that audience design places cognitive demands on the speaker, so speakers are more likely to engage in audience design when their cognitive load is lower (Horton & Gerrig, 2005; Roche, Dale, & Kreuz, submitted). Audience design may also be implemented when communicative situation calls for it (Galati & Brennan, 2010), rather than automicatically. Roche et al. (submitted) provide evidence that if speakers have time, they seek audience design strategies in communication. They suggest that, given time, speakers determine a strategy through monitoring and adjustment, a reconciliation of Galati and Brennan (2010)'s one-bit model and Horton and Keysar (1996)'s Monitoring and Adjustment model.

Our experiment involves a video description task for the speaker and a picture choice task for the listener. Then, the speaker and listener switched roles. If a speaker learns how to produce communicatively suitable utterances in the context of our experiment, their utterances should become more informative with time. In order to determine, what experience such learning may be based on, we ask if the speaker has previous experience as a listener facing difficult comprehension choices, will she be a more informative speaker? Or, does the speaker's own experience in production affect rate of informativeness?

## Experiment

We investigated the use of referring expressions in an experimental setting with speakers of Yucatec Maya, an indigenous language of Mexico. Our experiment is a video description task (for the speaker) coupled with a picture choice task (for the listener) based on the sentences produced by the speaker. In this experiment, we analyze rates of target specification, in which the speaker uniquely described a contrasting feature that distinguished two reference objects and compared that to rates of over- and under-specification. We examine whether informativeness is influenced by audience design.

### Materials and Methods

We created video stimuli using the Poser 3D character animation software. We had four conditions: 1) No Contrast, in which neither the theme nor the recipient showed a visual contrast (see Figure 1), 2) Recipient Contrast, in which only the recipient showed a visual contrast (see Figure 2), 3) Theme Contrast, in which only the theme showed a visual contrast (see Figure 3), and 4) Both Contrast, in which both the recipient and theme showed a visual contrast (Figure 4).

There were 24 ditransitive target videos (see Figures 1 through 4) in addition to 24 transitive and 16 intransitive filler videos. There were 64 listener pictures, 18 of which involved a minimal visual contrast which would require a description from the speaker that uniquely identified a contrasting feature in order for the listener to chose the correct picture (see Figure 5). The reminder did not involve a minimal contrast (see Figure 6).

Figure 1: No contrast - *Carlos gave the the donut to the man.*

Figure 2: Recipient contrast - *Carlos gave the donut to the man in the white shirt.*

**Participants** Thirty native speakers of Yucatec Maya took part in a communicative task at the Universidad del Oriente in Valladolid, Mexico and were compensated fifty Mexican pesos (about 5 U.S. dollars) for their participation. The experiment lasted no longer than one hour. Participants were recruited in pairs by asking them to bring a friend. There were nineteen females and eleven males (9 female-female pairs, 5 male-male pairs and one female-male pair). All participants were bilingual in Spanish and Yucatec Maya.

**Procedure and stimuli** Speakers were instructed to watch videos on a laptop screen and describe them using a single sentence in Yucatec. This video was not visible to the listener. Listeners saw a card with two pictures, one framed in red, the other green (see Figures 5 and 6), only on of which matched with the video. In front of the listener, there was a red and a green pile (indicated by a sheet of paper in the respective color). Listeners had to select the correct picture corresponding to the speakers description of the video by putting the card face down on the color pile that corresponded to the frame color of the correct picture. During the practice trials, the listener was told to show the speaker his or her cards so that the speaker would be aware of the information required to chose the correct picture, but participants were told that during the actual experiment the speaker would not see the pictures. Pictures were shown to the listener in such a way that a small blind obscured the line of vision between the speaker and the pictures but not between the speaker and the listener. Participants were asked to avoid using descriptions like "to

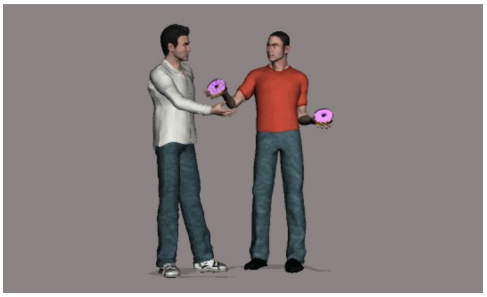

Figure 3: Theme contrast - *Carlos gave the the donut with the sprinkles to the man.*

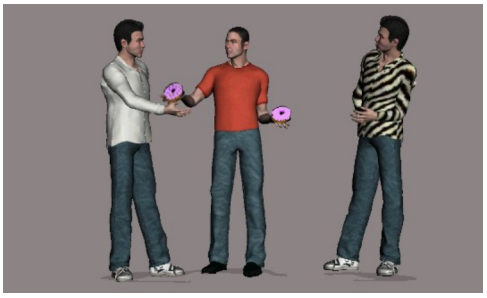

Figure 4: Both contrast - *Carlos gave the donut with the sprinkles to the man in the white shirt.*

Figure 5: Minimal contrast picture pair

Figure 6: No contrast picture pair

the left of" or "to the right of" because the pictures shown to the interlocutor might display referents in a different spatial order than shown to the speaker. Indeed, speakers saw such mirrored target pictures during the practice trials. Also, participants were informed that they would switch roles half way through the experiment and that the experimenter would indicate when it was time to switch roles and assist in the switching of roles. Participants had four practice trials at the beginning of the experiment and again after switching roles.

**Coding and analysis** The data were coded by the first author and checked against transcriptions by two native speakers of Yucatec. We looked at modifications that consisted of more than just one adjective, such as "small" or "pretty" unless it uniquely identify a theme or recipient, such as "the blue ball" versus the contrasting "white ball." We also included all types of modificational structures available in Yucatec Maya. We included afterthoughts in the analyses, such as "the boy...the one with the white shirt." Responses were scored as target specifications if there was a modification which identified the contrasting feature present in the video. Responses were considered under-specified if there was a visual contrast in the video but the speaker did not provide a description that identified the contrasting feature. Responses were scored as over-specified if there was a modification describing some feature for which there was no visual contrast in the video. We analyzed theme and recipient modifications for level of specification, under, over, or perfectly specified. Twenty-one percent of responses (164 out of 768) were excluded for being unintelligible, in Spanish, interrupted, or lacking all three arguments (agent, theme and recipient).

## Results and Discussion

### Informativeness

The presence of a visual contrast in the video significantly contributed to the likelihood of modification that uniquely identified the referent for themes ($\chi^2(3)=33$, $p < 0.001$) and recipients ($\chi^2(3)=111$, $p < 0.001$). Themes were under-specified more often than recipients (see Figure 7 for rates of target, under- and over-specification for themes and Figure 8 for recipients). In addition, themes were over-specified less often than recipients.

Speakers produced modifications that identified the contrasting feature for themes at a rate of 12 percent and for recipients at a rate of 41 percent. The rate of target modification only takes into account the relevant conditions in which there was a contrasting feature between two theme objects (The Both and Theme conditions for theme objects, and the Both and Recipient conditions for recipient objects). Speakers were over-informative (producing a modification of a reference object when there was no visual contrast present in the video) at a rate of about 6 percent for themes and 21 percent for recipients. Speakers were under-informative (failing to produce a unique description of a visual contrast differentiating two reference objects) in 88 percent of relevant cases for themes and in 58 percent of relevant cases for recipients.

### Listener comprehension

Overall, the comprehension accuracy was very high. In the No Contrast picture comprehension condition, where there

was not a minimal contrasting visual feature between two referents, participants chose the correct picture at a rate of 94%, which is expected given that the correct picture did not hinge upon the speaker's using a modification that uniquely identified a contrasting visual feature. In the Contrast picture comprehension condition, in which the correct choice of picture depended on a unique description of a theme or recipient (or both) on the part of the speaker, comprehenders chose the correct picture in 82% of cases. As would be expected, under-specified recipients greatly contributed to incorrect comprehension ($p < 0.002$). The effects of under-specification and Contrast picture did not interact.

### Learning Audience Design

Next, we investigated whether there is evidence for audience design changing rates of under- and over-specification. We investigated two hypotheses. First, it is possible that a speaker who plays the role of listener in the first part of the experiment will produce fewer non-target specifications when it's her turn to speak. This would be expected if speakers integrate their experience as comprehenders into their productions. Second, it is possible that experience with the production of referring expressions helps to find viable strategies that facilitate successful communication. To test these two hypotheses, we conducted four mixed logit analyses predicting the log-odds of (a) under-specification and (b) over-specification of theme and recipient expressions based on the (log-transformed) trial order within each half of the experiment, whether the production came from the first or second speaker, as well as the interaction of these two factors. The models also contained the maximal random effect structure justified by the data.

We found no effect of whether the speaker had first experienced the task as a comprehender (first and second speakers produced approximately the same rate of under- and over-specifications, but Figures 7 and 8 show that for themes and recipients, the second speaker produced slightly more target descriptions and slightly fewer under-specifications. The second speaker produced slightly fewer over-specifications for themes but over ten percent more over-specifications for recipients.

Though there was not a significant effect of speaker experience as a comprehender, we did find a significant main effect of trial order on the log-odds of under-specification for both themes ($p < 0.002$) and recipients ($p < 0.06$): under-specification became less likely the more trials the speaker had already produced (see Figure 9). Over-specification was not affected by trial order. The interaction of the two effects affected neither under- nor over-specification. In short, while listening to twenty-four production items by the partner does not affect what referential form speakers preferred, experiencing just a few production trials as a speaker seem to be sufficient to partially at least adapt to the task (by avoiding under-specification).

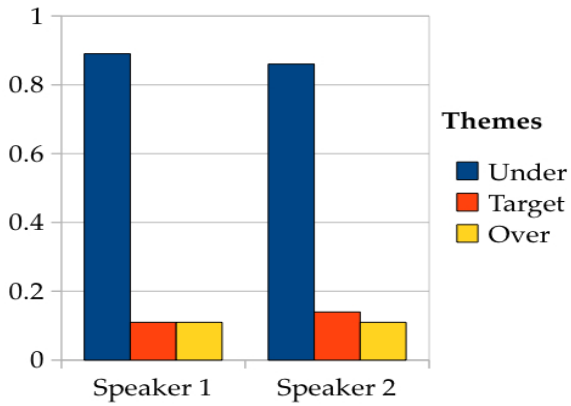


Figure 7: Proportion of under- correct and over-specification for theme descriptions

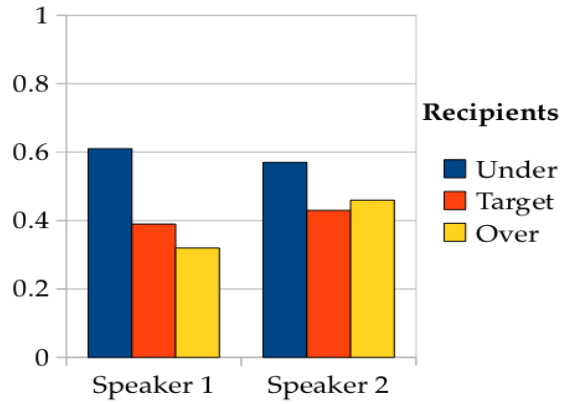


Figure 8: Proportion of under- correct and over-specification for recipient descriptions

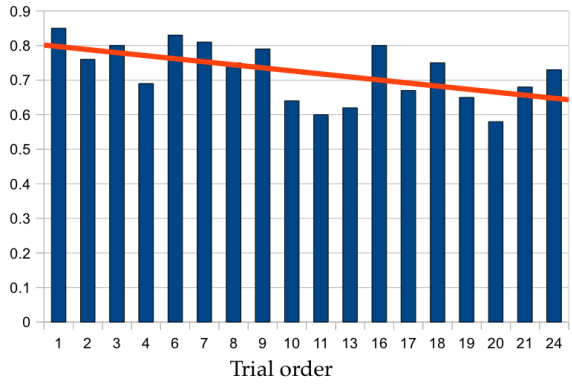


Figure 9: Rate of under-informativeness by trial order

## General Discussion

### Informativeness

We examined rates of target specification as well as over- and under-specification of themes and recipients of ditransi-

tive actions. Our Yucatec participants over-specified when referring to human recipients at about the same rate (33%) as speakers of other languages (28 to 50%) (Deutsch & Pechmann, 1982; Maes, Arts, & Noordman, 2004; Nadig & Sedivy, 2002; Pechmann, 1989). This finding expands the empirical coverage of psycholinguistic theories, showing that this phenomenon is not particular to the more common languages of inquiry. With regard to lower rates of over-specification of themes, however, our Yucatec Maya-speaking participants over-specified at a much lower rate (under 10%). It is possible that speakers were not seeing the contrasting features of themes, as these features were less visually salient. Non-linguistic cognitive factors such as visual salience have been shown to be influential in how informative speakers are (Davies & Katsos, 2009). We might also attribute this result to a difference in animacy (themes were inanimate, while recipients were human) or thematic or grammatical role.

### Audience design as learning

We found no effect of the speaker's previous experience as a comprehender on rate of informativeness in our experiment, contrary to Haywood, Pickering, and Branigan (2005)'s findings on ambiguity avoidance and Guhe and Bard (2008)'s findings on the over-specification of color features. It is possible that this difference between our results and those observed in Haywood et al. (2005) is due to adjustment to the situational communicative needs (Galati & Brennan, 2010), though in our experiment, participants switched roles after completing an entire list, while Haywood et al. (2005)'s participants switched roles after each trial).

We find a small but significant effect of the speaker's own experience on the rate of informativeness. Under-informativeness decreased as the speaker proceeded through the experiment trial-by-trial. As our speakers adjusted to the particular demands of the task, their utterances became more informative throughout the experiment.

Our results support audience design as learning, based on the speaker's experience with the particular experimental situation. The relatively small observed effect size may be due to the fact that our experiment was less interactive than experiments employing the referential communication paradigm (Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1964, 1966). The extent to which addressees are interactive in the communicative situation affects the extent to which speakers design their utterances for addressees (Lockridge & Brennan, 2002; Roche et al., submitted). Addressees who are silent or imaginary, versus addressees who provide feedback affects the extent to which speakers engage in audience design (Schober, 1993). We found our speakers to engage in audience design even though our addressees were silent.

Since, to the best of our knowledge, previous experiments have not directly contrasted learning of audience design based on production- vs. comprehension experience, it is also possible that production-based learning shows bigger effects and that the failure to observe listening-based learning in our experiment is due to a lack of power. This possibility requires further work comparing production- vs. listening-based learning of audience design.

In conclusion, our results support the idea that audience design may be cognitively demanding for the speaker (Rayner, Carlson, & Frazier, 1983; Horton & Gerrig, 2005; Roche et al., submitted) and may not be implemented automatically. Audience design may appear once speakers adjust to the situational communicative needs (Galati & Brennan, 2010). Speakers learn to design utterances for their addressees through a trial-and-error process (which could be Monitoring and Adjustment, Horton and Keysar (1996)).

## Acknowledgments

This research was funded by grants from the National Science Foundation to T. F. Jaeger (NSF 0848385) and J. Bohnemeyer (NSF 0848298) and an Alfred P. Sloan Fellowship to T. F. Jaeger. We thank Carlos Gómez Gallo for stimulus creation and data collection. We thank the Universidad del Oriente in Valladolid, Mexico for their space and Yucatec speakers.

## References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*, 55–70.

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

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*, 145–204.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formation. *Cognition*, *21*, 47–67.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *102*, 137–167.

Brown, P., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472.

Chafe, W. L. (1976). Subject and topic. In C. N. Li (Ed.), *Givenness, contrastiveness, definiteness, subjects, topics and point of view* (pp. 25–56). New York: Academic Press.

Chafe, W. L. (1994). *Discourse, consciousness, and time*. Chicago: Chicago University Press.

Christianson, K., & Ferreira, F. (2005). Conceptual acces-

sibility and sentence production in a free word order language (odawa). *Cognition*, *98*, 105–135.

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

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

Davies, C., & Katsos, N. (2009). Are interlocutors as sensitive to over-informativeness as they are to under-informativeness? In *Proceedings of the 2009 meeting of the cognitive science society.*

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of a model of the listener. In D. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics* (pp. 105–129). Academic Press.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159–184.

Eckert, P. (2004). The meaning of style. In W. Chiang, E. Chun, L. Mahalingappa, & S. Mehus (Eds.), *Texas linguistics forum* (Vol. 47). Texas Linguistics Forum.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying 'that' is not saying 'that' at all. *Journal of Memory and Language*, *48*, 379–398.

Ferreira, V. S. (2008). Ambiguity, accessibility and a division of labor for communicative success. *Psychology of Learning and Motivation: Advances in Research and Theory*, *49*, 209–246.

Ferreira, V. S., & Dell, G. S. (2003). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in japanese. *Journal of Psycholinguistic Research*, *32*, 669–692.

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

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.

Gregory, M. L., Healy, A. F., & Jurafsky, D. (2001). *Common ground in production: Effects of mutual knowledge on word duration.* (Unpublished manuscript, University of Colorado at Boulder)

Guhe, M., & Bard, E. G. (2008). Adapting referring expressions to the task environment. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions. *Language*, *69*(2), 274–307.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, *16*(5), 362–366.

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

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

Hudson, S. B., Tanenhaus, M. K., & Dell, G. S. (1986). The effect of the discourse center on the local coherence of a discourse. In *Proceedings of the 8th annual conference of the cognitive science society* (pp. 96–101). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hudson-D'Zmura, S. (1988). *The structure of discourse and anaphor resolution: The discourse center and the roles of nouns and pronouns*. Unpublished doctoral dissertation, University of Rochester.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric bias of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*, 46–50.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343–346.

Lindblom, B. (1990). Explaining variation: A sketch of the h and h theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 402–439). Dordrecht, the Netherlands: Kluwer Academic.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, *9*(3), 550–557.

Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes*, *37*(2), 117–144.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89–110.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in english and spanish. *Journal of Memory and Language*, *42*, 168–182.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence production. *Journal of Vebral Learning and Verbal Behavior*, *22*, 358–374.

Roche, J. M., Dale, R., & Kreuz, R. J. (submitted). *Don't rush the navigator: Disambiguation strategies require cognitive flexibility.*

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