The Role of Common Ground in Audience Design: Beyond an All or Nothing Story

Flora Vanlangendonck (flora.vanlangendonck@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

Roel Willems (roel.willems@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Laura Menenti (laura.menenti@mpi.nl)
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Peter Hagoort (peter.hagoort@mpi.nl)
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Abstract

We investigated when and how privileged information affects the production of referring expressions. Pairs of participants played a referential communication game in which speakers described target objects for their addressees. On critical trials, speakers saw competitor objects that were occluded from the addressee's view. We found that speakers already distinguish between shared and privileged information during the early stages of planning their referring expression. We found no evidence that tailoring referring expressions for the addressee requires additional planning time due to monitoring and adjustment. Our results support Constraint-Based Processing models of language production.

Keywords: audience design, common ground, eye-tracking, referential expressions, language production

Introduction

Speakers tailor their utterances for their addressees, a process known as audience design (Clark & Carlson, 1982). Key questions in research on audience design are when and how speakers take into account which knowledge is shared between speaker and addressee (*common ground*) and which information is only available to the speaker (*privileged ground*) (Clark & Marshall, 1978). According to Grice's Maxim of Quantity, speakers should make their contributions as informative as is required for the current purpose of an interaction (Grice, 1975). Information that is not shared between interlocutors and not informative should therefore not influence the production of referring expressions. However, speakers do not always successfully ignore privileged information during language production. Imagine a situation in which a speaker sees two bottles of different sizes, the smallest of which is occluded from their addressee's point of view. If the speaker asks the addressee to hand him the mutually visible bottle, he can either describe it using a size adjective (*the large bottle*) or a bare noun (*the bottle*). According to Grice's Maxim of Quantity, the speaker should not use a size-contrasting adjective, because the smaller bottle is occluded from the listener's point of view. Using a size adjective is therefore overinformative. However, speakers in this type of situation still regularly use a size adjective (Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008; Yoon, Koh & Brown-Schmidt, 2012). In many contexts, egocentrically produced referring expressions do not hinder communication, although they may temporarily confuse the addressee (Engelhardt, Bailey, & Ferreira, 2006). Yet, speakers also overspecify referring expressions when the additional information is uninformative (Wardlow Lane & Ferreira, 2008), and when they are instructed to conceal privileged information (Wardlow Lane, Groisman & Ferreira, 2006).

The question why speakers fail to ignore privileged information is still under debate, and two main theories have emerged to explain these results. According to Constraint-Based Processing models, speakers keep their addressee in mind from the earliest stages of utterance planning in a probabilistic, constraint-based way, resulting in early effects of common ground (Brennan & Hanna, 2009; Hanna et al., 2003; Horton & Gerrig, 2002; Tanenhaus & Trueswell, 1995). In this view, privileged and common ground information are partial constraints for language processing. When privileged information is salient, as in the example with the additional bottle above, speakers may fail to ignore privileged information. In contrast, the Monitoring and Adjustment theory proposes that keeping track of the distinction between shared and privileged information during utterance planning is resource-intensive and often not necessary (Horton & Keysar, 1996). According to this view, speakers initially design their utterances from their own egocentric perspective, and common ground only comes into play at a later stage when speakers monitor their utterance (Keysar, Barr & Horton, 1998). The Monitoring and Adjustment model thus predicts that speakers should initially not distinguish between privileged and common ground information during utterance planning. Monitoring and adjusting the initial utterance plan is considered

effortful and should require additional planning time (Horton & Gerrig, 1996). Speakers fail to ignore privileged information when the initial egocentric plan is not corrected.

In the current experiment, we investigated when and how shared and privileged visual information affect the production of referring expressions by means of eye-tracking and behavioural measures. In addition, we tested whether ignoring privileged information during utterance planning is resource-intensive, as reflected in longer speech planning times for trials with a relevant visual perspective difference. Finally, we were interested in whether speakers avoid egocentric descriptions when they threaten communicative success. In Wardlow Lane and Ferreira's experiment (2008), target objects were in privileged ground in certain filler trials, meaning that speakers could not systematically ignore privileged ground information. In the current experiment, we tested whether speakers are sensitive to the possibility of miscommunication when they are able to ignore privileged information during the whole task. We therefore introduced two types of conditions. In the advisable audience design condition, a failure to ignore privileged information was overinformative, but the intended referent was clear. In the obligatory audience design condition, a failure to ignore privileged information forced the addressee to guess which referent was intended.

Method

Participants

22 pairs of native Dutch speakers played a computerised version of a referential communication game (Keysar et al., 2000; Yoon et al., 2012). Participants did not know each other before the start of the experiment. Data from one pair were excluded from the analyses due to experimenter error and data from another pair were excluded because the speaker's responses could not be coded. The remaining participants consisted of 6 men and 34 women (mean age: 21.5 years). Half of the eye-tracking data from one pair were lost due to equipment malfunction.

Procedure

A coin toss was used to randomly assign participants to the roles of speaker and listener for the duration of the experiment. The speaker and listener were seated at separate monitors, separated by a screen to prevent them from seeing each other's monitor. Throughout the experiment, participants saw opposite sides of a 4 x 4 array containing objects of different sizes (Figure 1). Each array contained 3 closed slots on each player's side, allowing us to manipulate which objects were in common ground.

Participants completed 288 trials in total. On each trial, the speaker described a specific object in the array in a way that would allow the listener to select the correct object from the array. During the first phase of the trial (3000 ms), speakers and listeners each saw their side of the array to allow them to establish which objects were in common ground. Then the speaker was cued by means of a red circle around one of the objects, and gave a description (Figure 2). On the basis of the speakers' response, the listener selected an object by means of a mouse click. The cued object was always mutually visible. Speakers were instructed not to use descriptions referring to the position of the object in the array.

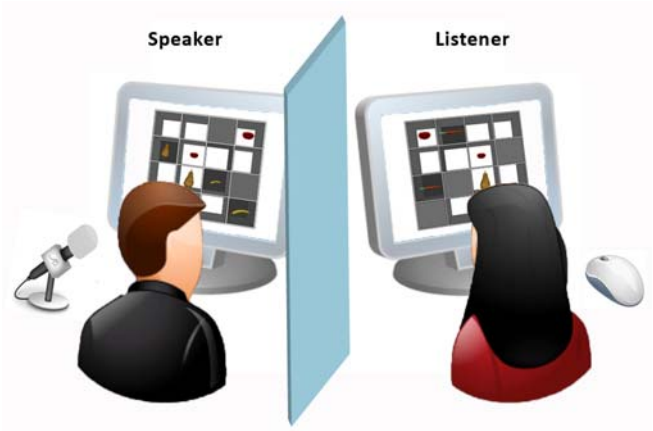

Figure 1: Overview of the set-up.

Before starting the experiment, participants practised the task together using a real array and real objects. They jointly placed the objects in the array, viewed each other's perspective, and performed the task. Then participants practised the task on the computer; during part of these test trials they were allowed to see both monitors and they could give feedback to each other.

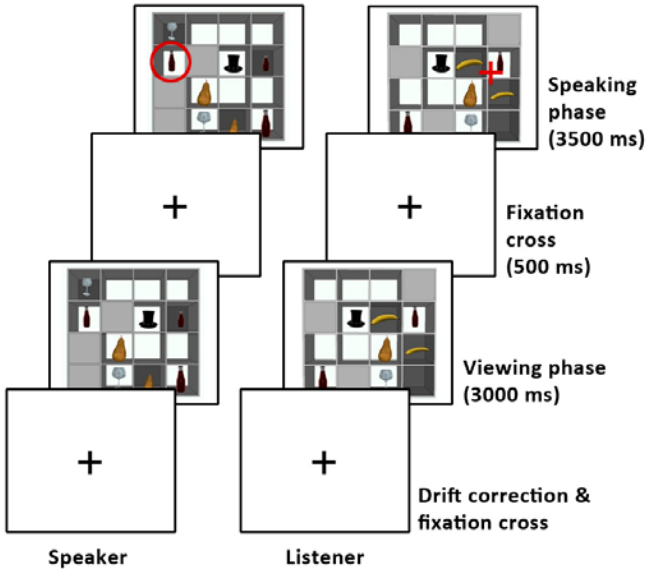

Figure 2: Trial sequence from the speaker and the listener's point of view.

Speakers' descriptions were recorded and their eye movements were tracked using a head-mounted EyeLink II eye-tracker (SR Research) with a 250 Hz sampling

frequency. We tracked both pupil and corneal reflection whenever possible. The distance between the speaker and the monitor was approximately 57 cm. We calibrated the eye-tracker at the start of the experiment, after a break halfway through the experiment and whenever deemed necessary. Drift correction was performed before each trial.

Materials

We manipulated the number, size and visibility of the relevant objects to create 6 conditions (Figure 3). In the audience design conditions, speakers saw an extra competitor object that the listener could not see. In the obligatory audience design condition, speakers saw 3 relevant identical objects of different sizes: one target object, one occluded competitor object and one mutually visible object. The target object was always the medium-sized object of the 3 objects. If speakers described this object from their own perspective, they would call it the *medium* object. In this case, their addressee had a 50 per cent chance of selecting the wrong object. On the other hand, if speakers considered the perspective of their addressee, they would ignore the occluded object and call the medium-sized object *small* or *large*. In the advisable audience design condition, speakers saw 2 relevant identical objects of different sizes: one target object and one occluded competitor object. Given that their addressee could see the target object but not the competitor object, speakers did not have to use a contrasting size adjective. However, unlike in the obligatory audience design condition, listeners were still expected to be able to select the correct object if the speaker did not consider the perspective difference.

We created two types of control conditions in which speakers and listeners saw the same number of relevant objects. In the linguistic control conditions, the occluded object was replaced by another, unrelated object. As a result, speakers saw one less relevant object in these conditions than in the audience design conditions. They were expected to produce the same utterance in these conditions as in the audience design conditions if they successfully ignored privileged information and adapted to their addressee's perspective. In other words, they were expected to produce a contrasting size adjective (*small/large*) in the obligatory linguistic control condition and a bare noun in the advisable linguistic control condition. In the visual control conditions, the object that was occluded in the audience design conditions was visible to both participants. As a result, speakers and listeners could both see all relevant objects. Speakers were expected to produce the same utterance in these conditions as in the audience design conditions if they did not successfully ignore privileged information. We expected speakers to describe the target object as *medium* in the obligatory linguistic control condition and as *small/large* in the advisable visual control condition.

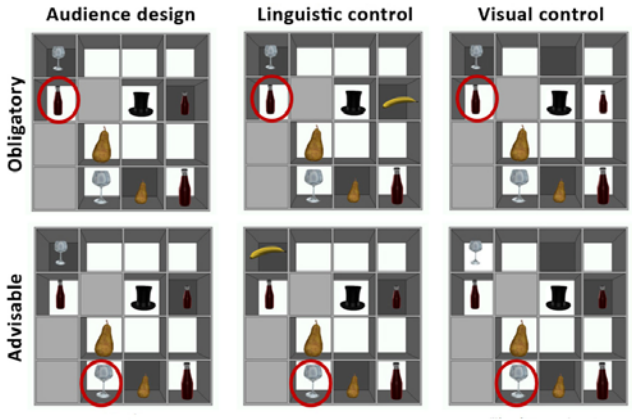

Figure 3: Overview of the 6 conditions used in the experiment.

We created 12 different virtual arrays. Each array was used in 24 trials in total, divided over 4 blocks of 6 trials. The arrays contained 6 to 8 table-top objects chosen from a total of 22 table top objects. Objects were selected from the Object Databank (courtesy of the Michael Tarr lab, Brown University, Providence, RI). Each object could appear in 4 different sizes to make sure that participants could not rely on absolute size. Depending on the condition, speakers saw 1, 2 or 3 relevant objects of the same type but of different sizes. The remaining objects were fillers that also appeared in sets of 1, 2 or 3 objects of the same type to make sure participants could not predict which objects would be relevant. We made sure speaker and listener always saw the same total number of objects in a trial by adding filler objects to the occluded slots if needed.

Results

We coded the sound files for adjective use. For the obligatory trials, we coded the use of bare nouns and *small/large* responses (*klein* and *groot* in Dutch). For the advisable trials, we coded the use of *small/large* and *medium* adjectives (*middelgroot* in Dutch). In addition, we coded errors, false starts, repairs, speech unrelated to the task, and responses exceeding the response interval of 3.5 s, and removed these trials from the dataset for analysis.

We analysed the results using linear mixed models (Baayen, Davidson & Bates, 2008) with a random intercept for subjects, a by-subjects random slope for condition, and condition and number of objects as fixed effects. We used the same model in every analysis, except that the family was set to reflect the distribution of the dependent measure: Poisson for count data (number of fixations per trial), binomial for binary data (adjective use) and Gaussian for continuous data (planning durations, speaking durations, and fixation durations). Obligatory and advisable data were analysed separately. The obligatory conditions contained more relevant objects than the advisable conditions, so we could not compare them directly. For each dependent measure, we ran planned contrasts comparing the audience design condition with each of the control conditions.

Adjective Use

We computed the percentage of each type of response for the different conditions to find out how often speakers failed to ignore privileged information when producing a referring expression. For the advisable conditions, we treated the use of a size adjective (*small* or *large*) as a binary dependent variable; for the obligatory trials, we treated the use of the adjective *medium* as a binary dependent variable.

Speakers adapted their language use to their addressee's perspective in the majority of audience design trials (Figure 3), yet they failed to ignore privileged information on all trails. In the obligatory audience design condition, speakers mainly produced *small/large* responses (90.89%), although they also used *medium* responses (9.11%). In the linguistic control condition, speakers mostly produced utterances that contained a *small/large* size adjective (97.45%), while they mainly produced *medium responses* (98.68%) in the visual control condition. Both the difference between the obligatory audience design condition and the linguistic control condition ($b = -3.87$, $SE = 0.52$, $p < 0.001$) and the difference between the obligatory audience design condition and the visual control condition were significant ($b = 8.59$, $SE = 0.71$, $p < 0.001$).

In the advisable audience design condition, speakers mostly produced bare nouns (81.61%), although they also produced *small/large* responses (18.39%). In the linguistic control condition, speakers mainly produced bare nouns (88.72%), and in the visual control condition they predominantly used *small/large* responses (98.91%). Both the difference between the advisable audience design condition and the linguistic control condition ($b = -1.64$, $SE = 0.28$, $p < 0.001$) and the difference between the advisable audience design condition and the visual control condition were significant ($b = 6.60$, $SE = 0.60$, $p < 0.001$).

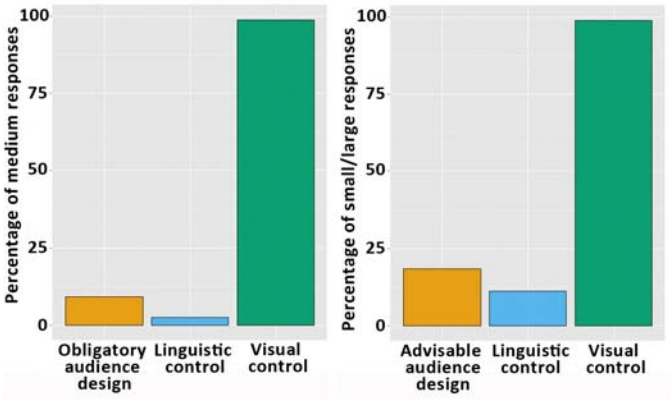

Figure 3: The percentages of speakers' *small/large* and *medium* responses for the obligatory and advisable data.

Planning and Speaking Durations

We computed speakers' planning durations (from cue onset to speech onset) based on the sound recordings to test whether ignoring privileged information during utterance planning is resource-intensive. We found no significant differences in planning durations between the obligatory audience design condition and the linguistic control condition ($b = -0.027$, $SE = 0.025$, $p = 0.42$), nor between the obligatory audience design condition and the visual control condition ($b = -0.0053$, $SE = 0.028$, $p = 0.97$). Similarly, we did not find significant differences between the advisable audience design condition and the linguistic control condition (audience design $b = -0.013$, $SE = 0.015$, $p = 0.58$), nor between the advisable audience design condition and the visual control condition ($b = -0.017$, $SE = 0.020$, $p = 0.58$).

We also calculated speaking durations (from speech onset to speech offset). If ignoring privileged information is a resource-intensive process, this might also lead to longer speaking durations. No significant difference in speaking time was found between the obligatory audience design condition and the linguistic control condition ($b = -0.017$, $SE = 0.0085$, $p = 0.078$). Speaking durations were shorter in the obligatory audience design condition than in the visual control condition ($b = 0.085$, $SE = 0.014$, $p < 0.001$), most likely because the Dutch word *middelgroot* has more syllables than the other size adjectives. We found longer speaking durations in the advisable audience design condition compared to the linguistic control condition ($b = -0.027$, $SE = 0.0090$, $p < 0.01$), and shorter speaking durations in the advisable audience design condition than in the visual control condition ($b = 0.22$, $SE = 0.018$, $p < 0.001$). These differences seem to stem from the use of adjectives compared to bare nouns. In the visual control condition, speakers systematically produced size adjectives. When we removed the trials with a size adjective in the advisable audience design condition, the significant difference between the advisable audience design condition and the linguistic control condition disappeared.

Number of Fixations per Trial

In order to measure when privileged information affects the production of referring expressions, we computed the mean number of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions. We focused our analysis specifically on the planning duration window. For both the obligatory and the advisable data, we found that speakers fixated an occluded competitor object (audience design conditions) more than an unrelated control object (linguistic control), but less than a mutually visible competitor object (visual control). Tables 1 and 2 give an overview of the number of fixations per trial.

We found significantly more fixations per trial in the obligatory audience design condition than in the linguistic control condition ($b = -1.57$, $SE = 0.15$, $p < 0.001$). We found significantly less fixations per trial in the obligatory audience design condition than in the visual control condition ($b = 0.88$, $SE = 0.12$, $p < 0.001$). To rule out that these differences were driven by the small percentage of obligatory control trials on which speakers responded with a *medium* response, we ran the same analysis again including

only trials in which speakers' responses took into account the listener's perspective (i.e., a medium response in the visual control condition, a small/large response in the audience design and linguistic control conditions). The differences remained significant in this analysis. Across the obligatory conditions, the mean number of fixations per trial is relatively low, even when the competitor object was relevant (visual control condition). Some participants reported that they paid special attention to objects that appeared in triplets during the first half of the trials, which may have allowed them to plan their utterance after the cue without additional fixations on the other objects.

We found a significant difference in the mean number of fixations per trial between the advisable audience design condition and the linguistic control condition ($b = -1.46$, $SE = 0.13$, $p < 0.001$) and between the advisable audience design condition and the visual control condition ($b = 0.31$, $SE = 0.060$, $p < 0.001$). This pattern remained even when we only included the trials on which speakers' responses took into account the listener's perspective (i.e., an adjective in the visual control condition, a bare noun in the advisable audience design and linguistic control conditions).

Table 1: Fixation results of the obligatory data. Standard deviations are in parentheses.

Measure	Obligatory audience design	Linguistic control	Visual control
Number of fixations	0.30 (0.55)	0.07 (0.28)	0.65 (0.68)
Total fixation time	164.80 (97.06)	146.86 (74.86)	180.29 (96.54)

Table 2: Fixation results of the advisable data. Standard deviations are in parentheses.

Measure	Advisable audience design	Linguistic control	Visual control
Number of fixations	0.65 (0.69)	0.17 (0.39)	0.87 (0.72)
Total fixation time	198.58 (103.32)	135.84 (83.34)	222.37 (114.08)

Total Fixation Duration per Trial

We computed the total duration of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions. We limited the analysis to fixations during the planning duration window. We treated trials without fixations on the occluded object as missing data. Speakers fixated the unrelated occluded object in the linguistic control condition in only 6.49% of obligatory linguistic control condition trials, and on 16.08% of advisable linguistic control trials. Given that there were so few of these trials, we did not contrast the audience design conditions and the linguistic control conditions for this dependent variable. Tables 1 and 2 give an overview of the total fixation duration per trial.

Speakers looked longer at mutually visible objects (visual control conditions) than at objects that were occluded from the listener's point of view (audience design conditions). We found a significant difference in total fixation duration between the obligatory audience design condition and the visual control condition ($b = 30.02$, $SE = 9.46$, $p < 0.01$). Similarly, speakers spent less time looking at the occluded competitor object in the advisable audience design condition than at the mutually visible object in the visual control condition ($b = 22.48$, $SE = 7.34$, $p < 0.01$).

Discussion

In this experiment, we investigated the effect of privileged information on the production of referring expressions. Speakers mainly produced referring expressions that took into account their addressee's visual perspective. However, the availability of privileged information also led to the production of utterances that did not take this perspective difference into account. These finding are in line with previous studies using similar paradigms (Yoon et al., 2012; Wardlow Lane et al., 2006). Interestingly, speakers even failed to completely ignore privileged information when it harmed communication (i.e., the obligatory audience design condition). In principle, speakers could fully ignore occluded objects throughout the experiment, yet even threats to communicative success did not prevent the occasional interference of privileged information.

In addition to studying the form of speakers' referring expressions, we collected planning durations and eye-tracking data to address when and how privileged information affects the production of referring expressions. Speakers fixated occluded competitor objects less than relevant competitor objects both in terms of the number of fixations and the total fixation duration. This suggests that speakers did not initially treat privileged information in an egocentric manner as predicted by the Monitoring and Adjustment hypothesis. Instead, they seemed to use the information that was available to them in the first half of the trials to distinguish between common and privileged ground when they were planning their referring expression. However, speakers did not fully ignore privileged information. We found more fixations on the occluded target objects than on unrelated occluded objects. This shows that speakers' general success at producing utterances that took into account their addressee's perspective was not the result of fully ignoring all occluded objects. Although privileged information was never relevant in our experiment, speakers could not avoid paying attention to objects in privileged ground.

To test the prediction of the Monitoring and Adjustment theory that audience design should lead to additional planning time (Horton & Keysar, 1996), we computed and compared planning durations. However, unlike Horton & Keysar (1996), we did not find any significant differences in planning duration between the audience design conditions and the control conditions. We thus found no evidence that planning a referring expression in the presence of an

occluded competitor object is necessarily more resource-intensive.

Taken together, our results show that common ground does not function as a complete constraint on the production of referring expressions, but does exert early effects during utterance planning. Speakers tried to use the information that was available to them during the first half of each trial to restrict common ground when they were planning their utterance, as reflected in the lower number of fixations on the occluded object during planning in the second half of the trials. When speakers fixated the occluded object during planning, this often did not prevent them from tailoring their referring expression for their addressee. These results support Constraint-Based Processing models. In this view, common and privileged ground act as probabilistic constraints to guide language processing in combination with other constraints such as context (Brown-Schmidt & Hanna, 2011). Given the lack of differences in planning durations and speakers' relative success at tailoring their referring expressions for their addressee, the process of weighing the available shared and privileged information appears to be a relatively efficient process.

Conclusion

In line with previous studies using similar paradigms (Yoon et al., 2012; Wardlow Lane et al., 2006), we found that speakers cannot completely ignore privileged information during language production and that the availability of privileged information can lead to the production of utterances that do not take the addressee's visual perspective into account. However, we found no evidence that adapting to your addressee' visual perspective requires additional planning time, suggesting that audience design is not necessarily effortful. Combined, these results suggest that speakers can efficiently constrain the available privileged and shared information during language production. These findings support Constraint-Based Processing models that predict that common and privileged information is incorporated into language processing in a probabilistic fashion. Our results show that common ground does not exert an all-or-nothing influence on language production, but influences the production process already during utterance planning as a partial constraint.

References

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18-40.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action), 1*, 274–291.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialog and Discourse, 2*, 11-33.

Clark, H. H. & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*, 332-373.

Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical Issues In Natural Language Processing* (Vol. 2). New York: Association for Computing Machinery.

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, 54*, 554-573.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Arts*. New York: Academic Press, 1975.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43-61.

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

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

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*(2), 46-50.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38.

Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech and Language*. San Diego, CA: Academic Press.

Wardlow Lane, L., Groisman, M., Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science, 17*, 273–277.

Wardlow Lane, L. & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481.

Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*, 699-707.