How big is the BFG? The impact of redundant size adjectives on size perception

Emiel Krahmer (E.J.Krahmer@uvt.nl)
Tilburg center for Cognition and Communication (TiCC), School of Humanities,
Tilburg University, The Netherlands

Marret K. Noordewier (M.K.Noordewier@fsw.leidenuniv.nl)
Faculty of Social and Behavioral Sciences, Social and Organizational Psychology,
Leiden University, The Netherlands

Martijn Goudbeek (M.B.Goudbeek@uvt.nl)
Tilburg center for Cognition and Communication (TiCC), School of Humanities,
Tilburg University, The Netherlands

Ruud Koolen (R.M.F.Koolen@uvt.nl)
Tilburg center for Cognition and Communication (TiCC), School of Humanities,
Tilburg University, The Netherlands

Abstract

In two experiments we study how redundant size modifiers influence the perceived size of objects. We show that when objects are referred to with overspecified descriptions (for example, using a description like "the large red chair" in a situation where all chairs are equally large but different in color), participants subsequently estimate the object to be larger than when objects are referred to using minimally distinguishing descriptions (e.g., "the red chair"). In Experiment 1, we show this effect with adult language users and different kinds of size modifiers. In Experiment 2, the same effect is shown for children of two different age groups (7- and 10-year olds), and for different kinds of visual size contrasts. Interestingly, we observe an inversely proportional relation between the age of our child participants and the difference in size estimates for minimal and overspecified descriptions, suggesting that language users gradually become better at avoiding false pragmatic inferences from redundant adjectives as they grow older.

Keywords: Reference, Overspecification, Language Development, Conversational Implicature

Introduction

Arguably, referring to a giant as "big" is somewhat excessive, certainly when there is only one giant in sight. But would calling a giant "big" and "friendly" (as Roald Dahl, 1982, does in his well-known children's novel *The BFG*; short for Big Friendly Giant), nevertheless have an impact on the perceived size (or friendliness) of said giant? And would this effect be the same for younger children as for older ones or even for adults? These are essentially the questions we address in this paper.

Background

Speakers frequently produce definite descriptions such as "the big friendly giant", "the red chair" and "the large ball", since they allow them to link their utterances to the physical world surrounding them. One central problem that a speaker has to solve when planning such a referring expression is to decide which properties to include in the reference. A chair can be red, but also large, plush, modern, with or without cushions and armrests, cheap or expensive, etc. So which properties to select? A successful reference includes sufficiently many properties to allow the addressee to determine which chair the speaker has in mind, but not too many, as Dale and Reiter (1995) propose in their computational interpretation of Grice's (1975) maxims for reference production.

One *prima facie* plausible option would be to opt for the smallest set of properties that distinguish the target object from the other objects in the context (Dale 1989). However, it has been repeatedly found that this is not necessarily what speakers do (e.g., Olson, 1970; Pechmann, 1989; Belke & Meyer, 2002; Engelhardt et al., 2006; Koolen et al., 2011). In many cases, speakers produce *overspecified* descriptions, which contain one or more redundant modifiers. For example, they produce a description such as "the large red chair", in a situation where "the red chair" would have been sufficient to single out the target. A number of speaker-internal factors have been shown to influence the likelihood of speakers producing an overspecified description, ranging from the pressures of incremental speech production (speakers may start producing a referring expression before scanning of a visual scene is complete; Pechmann, 1989) to scene complexity (more overspecification in complex visual scenes; Koolen et al., 2012) and conversational setting (more overspecification when misunderstandings are costly; Arts et al., 2011), suggesting that overspecification does not have a single distinct cause.

However, in this paper we focus on the impact of overspecification on language *understanding*, and here the picture is less clear. Some researchers have suggested that overspecification can help addressees with identifying an intended target (e.g., Arts et al., 2011; Paraboni et al., 2007), while others argued that they slow down identification (e.g., Engelhardt et al., 2006; 2011). Importantly, all these studies focus only on object identification. In this paper, we argue that overspecification in referring expressions may also have other important side effects for addressees.

A prominent view in language understanding is based on the assumption that "utterances convey only relevant information" (Frank & Goodman, 2012). This assumption can be traced back at least to the work of Grice (1975), who postulates among other things that speakers should not make their contribution more informative than is required (this is half of his well-known Maxim of Quantity). A speaker that violates ("flouts") this maxim, by providing more information than needed, thereby triggers a conversational implicature, suggesting to the listener that the additional material is meaningful after all. Imagine, for instance, that a speaker tells you to "sit by *the newly-painted table*" (Dale & Reiter, 1995), while there is only one table in the room. In that case, you may think the modifier "newly-painted" is redundant (since it is more informative than required for the purpose of identification), and this might cause you to infer the conversational implicature, intended by the speaker, that it is best not to rest your arms on this table in order to keep your clothes unstained.

However, one can also think of situations where an addressee may reason that the redundant information is somehow relevant, even when the speaker did *not* intend it in this way (Grice would call this a *false* conversational implicature). After all, as we argued above, speakers may overspecify for a variety of reasons. Our first hypothesis therefore is that if an object is described redundantly as "large" or "small" this will influence how the size of this object is perceived. Redundantly referring to a target object as large (small) may cause people to perceive or remember the target as larger (smaller) than when such a redundant size modifier is not included in a description. Even though we focus on redundant size adjectives here, we conjectured that other redundant adjectives (e.g., referring to color) could have similar effects (a possibility we discuss below).

Moreover, it has been argued that children are more likely to derive false conversational implicatures than adults (Siegal & Surian, 2004). On the one hand, we know from earlier research that children have a general tendency to regularly produce underspecified or ambiguous referring expressions until they are about seven years old (Deutsch & Pechmann, 1982; Matthews, Lieven, & Tomasello, 2007), and before that age only marginally benefit from redundant information in target identification (Sonnenschein, 1982; Ackerman, Szymansi & Silver, 1990; Davies & Katsos, 2010). Indeed, one could argue that it requires relatively sophisticated pragmatic reasoning to understand the implications of redundant information in descriptions. Therefore our second hypothesis is that children are more susceptible to redundant size modifiers than adults when making size estimates. Given that earlier work suggests that the relevant pragmatic reasoning is under development until children are about 7 years old, we test this both with child participants of on average 7 years (Group 3 in the Dutch elementary school system) and 10 years old (Group 6) in the experiment.

The current studies

We test these two hypotheses in two experiments (one with adults, one with children of two age groups), which rely on the same basic idea: participants hear descriptions referring to objects in a visual scene. Descriptions can either be minimally specified or overspecified (containing a redundant size modifier). After participants have processed a description, the objects in the visual scene disappear from view, and participants are asked to indicate how large they think the target object (which no longer is visible) was.

Experiment 1: Adults

Method

Participants Participants were 68 undergraduate students from Tilburg University (49 female) who participated for partial course credits. Their mean age was 21.5 years (SD = 2.4). All were native speakers of Dutch, the language of the experiment.

Figure 1: Example scene consisting of two same size chairs, with different colors (may not be visible in a black and white print). In the minimal condition, the target would be referred to as "the red chair", while in the overspecified condition it would be "the large red chair".

Materials Stimuli were created using pictures of furniture items from the Object Databank, created by Michael Tarr and colleagues, and often used in research on reference (e.g., van Deemter, Gatt, van der Sluis & Power, 2012). Three furniture items were selected for the current experiment (chair, couch, desk), and manipulated for color (either red or blue) and size (either large or small). Each stimulus consisted of two objects, one target (the object being referred to) and one distractor. There were 12 different targets (3 object types x 2 colors x 2 sizes), and the left-right position of the target with respect to the distractor was counterbalanced. In the critical stimuli, the distractor was always of the same type and size as the target and only differed in color (see Figure 1 for an example). Each target was referred to once with a minimal description (e.g., "the red chair") and once with an overspecified description containing a redundant size adjective (e.g., "the large red chair" or "the small red chair", depending on the size of the target). This created a total of 12 x 2 = 24 critical trials. The experiment also contained 24 filler trials, in which the visual objects consisted of different types, sizes and colors.

Procedure During the experiment, individual participants were seated in front of a computer screen, on which pairs of objects were presented as in Figure 1, together with a pre-recorded spoken description (e.g., "the [large] red chair"), produced by a female speaker with neutral intonation (i.e., with nuclear stress on the noun and no pitch accent on the adjective) and presented to participants over headphones. After a fixed interval, both objects disappeared from the screen and a horizontal slider appeared, together with the question "How large was the _____?", where the gap in this question was filled by the type of the target (e.g., "chair") The slider had the shape of an elongated, isosceles triangle with the tip ("small") on the left- and the base ("large") on the right-hand side (see Figure 2). Upon appearance, the slider handle was positioned in the middle; the handle had to be moved before the participant could proceed to the next stimulus. For analysis, the position after being set by the participant was mapped to a score between 0 and 100 (with higher number indicating larger size estimates).

In addition to the size question, participants were also asked to indicate the color of each object referred to on a one-dimensional saturation scale, ranging from lighter to darker, again with the handle initially positioned in the middle, on the assumption that a redundant mention of color (like "red") would cause participants to perceive an object as "redder" than when color was not mentioned in a description. However, no reliable effects of redundant color adjectives were found, and we will not describe the results of this measure further. In the general discussion we do return to this issue.

Experiment 1 had a within-participants design: all participants produced a size estimate for all targets. Stimuli were presented in a random order. Before the actual experiment started, a three trial training session (with a fan as target object type) was presented, to make participants familiar with the experimental set-up. After the training session there was no further interaction between participants and experimenter.

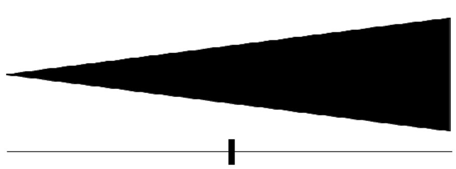

Figure 2: Slider used in Experiment 1 for size estimates.

Statistical analysis To test for significance, we conducted a 2 x 2 repeated measures Analysis of Variance (ANOVA), with Size (levels: large, small) and Description (levels: minimal, overspecified) as independent variables and average size estimate as the dependent variable.

Results and discussion

Figure 3 summarizes the results. First of all, a main effect of Size was found, $F(1,67) = 461.94$, $p < .001$, $\eta^2 = .87$. Large targets were estimated to be larger ($M = 60.45$, 95% CI = (57.83, 63.07)) than small ones ($M = 23.17$, 95% CI = (20, 26.34)). This serves as a manipulation check and indicates that the slider worked exactly as intended. In addition, a main effect of Description was found, $F(1,67) = 5.30$, $p < .05$, $\eta^2 = .07$. Targets that were referred to using an overspecified description were estimated to be larger than targets that were referred to using minimal descriptions. Importantly, this main effect was qualified by an interaction between Size and Description, $F(1,67) = 15.16$, $p < .001$, $\eta^2 = .18$. This interaction can be explained by inspection of Figure 3: large targets that are referred to redundantly are estimated to be larger ($M = 62.81$, 95% CI = (59.85, 65.76) than ones that are referred to minimally ($M = 58.1$, 95% CI = (55.34, 60.85)), while small targets that are referred to redundantly are estimated to be smaller ($M = 22.10$, 95% CI = (18.88, 25.33)) than ones that are referred to minimally ($M = 24.24$, 95% CI = (20.86, 27.62)).

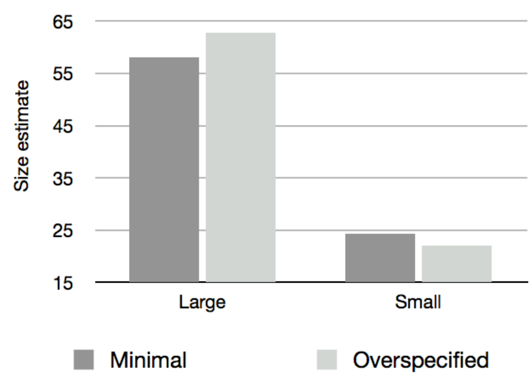

Figure 3: Mean size estimates in millimeters (range 1-100) for large and small objects for minimally specified descriptions (e.g., "the red chair") or overspecified ones ("the large/small red chair").

Experiment 1 clearly showed that adults are sensitive to redundant size modifiers in distinguishing descriptions. In Experiment 2, we conduct a comparable experiment with child language users in two different age categories, to see whether younger and older children are similarly sensitive to redundant modifiers. In addition, in this experiment we also vary the visual size of the target, to see whether size differences between target and distractor influence any effects of redundant size modifiers.

Experiment 2: Children

Method

Participants Sixty normally developing children were included in the study, in two age groups: 30 younger children (13 girls, 17 boys), with an average age of 7.1 years (range: 6.6-8.3), all in Group 3 of the Dutch elementary school system; and 30 older ones (15 girls, 15 boys), with an average age of 10.2 years (range: 8.7-11.5), all in Group 6. The children participating in this experiment were native speakers of Dutch, the language of the experiment. All children came from the Mgr. Zwijsenschool in Kerkdriel (Gelderland, The Netherlands). Parental consent for participation in the experiment was obtained for all children prior to the experiment.

Materials Pictures of eight different photorealistic children's toys (football, teddy bear, train, slide, rubber duck, doll, boat, spin top) were used as targets in this experiment (see Figure 4 for two representative examples). Each target was presented together with one distractor toy. Four different conditions were created for each of the eight targets by varying the size of the distractor (which could either be depicted as large as or smaller than the target) and by varying the reference to the target (which could either include a redundant size modifier or not), resulting in 8 x 4 = 32 critical trials. Note that in this experiment each target could uniquely be identified by its type ("the football"), so including a size adjective always resulted in an overspecified description. In addition, eight control trials were included, one for each target type, in which the target was combined with a smaller object of the same type (e.g., a small football), so that the size adjective in a description such as "the large football" was informative and not redundant. This allowed us to check whether any differences in size-estimates for *non*-redundant adjectives between age groups could be observed. In all 40 stimuli the left-right position of the target with respect to the distractor was counterbalanced.

Figure 4: Examples of visual stimuli used in Experiment 2, with two different children's toys. Again the target could either be referred to in a minimal way ("the ball") or an overspecified one ("the large ball").

Procedure Children performed the experiment individually, and were seated in front of a computer monitor in a quiet room in the school building. The procedure for younger and older children was exactly the same and went as follows: after a brief training session, in which the magnitude estimation scale was practiced, stimulus presentation started with a pair of toys presented on a white background for 4 seconds. After this a white screen appeared for 8 seconds, during which children were asked to answer a pre-recorded question "How large was _____?" The gap in this question was filled by a description of the target, which could either be minimal ("the football") or overspecified ("the large football"). For the audio recordings, a male adult speaker was used, who realized each question with a neutral intonation. Since the experiment was conducted in Dutch, this implies that the nuclear stress always occurred on the noun and the adjective was produced without a pitch accent.

Children were asked to indicate their size estimate on a magnitude estimation scale of 100 millimeters (consisting of a horizontal line without units of length added), with on the left-hand side a picture of the target reduced by a factor of 1.5, and on the right hand side the same picture enlarged by a factor of 1.5, in such a way that the real value was exactly in between (remember that the target figure was not visible to the child during the size estimation phase of the experiment, so children could not directly map the perceived size onto the scale). Children could indicate the estimates on paper, using a booklet that was positioned in front of them. After completing one trial, the next pair of toys appeared on the screen. The entire experiment lasted approximately 10 minutes. After the experiment, size estimates were manually measured in millimeters, with higher numbers indicating larger sizes. Measurements were done blind for condition.

Design and analyses The experiment had a 2 x 2 x 2 mixed design, with Description (minimal, overspecified) and Size (target and distractor equally large, target larger than distractor) as within-participant factors, Age group (younger, older) as a between-participant factor and average size estimate as the dependent variable. Tests for significance were conducted using a repeated measures ANOVA.

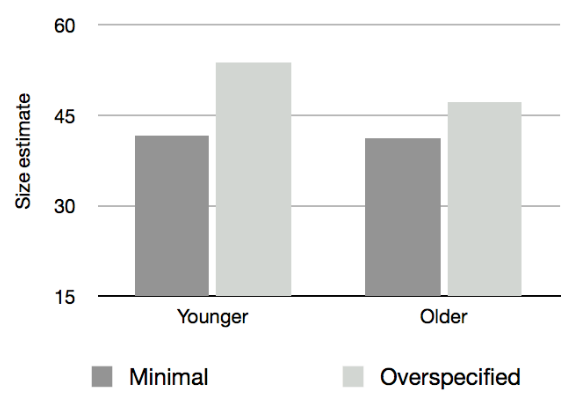

Figure 5: Mean size estimates in millimeters (range 1-100) of younger (avg. 7.1 years) and older (avg. 10.2 years) children for targets that were minimally specified (e.g., "the ball") or overspecified ("the large ball").

Results and discussion

The results showed that Size had a significant impact on children's size estimates. If the distractor was smaller than the target, children perceived the target as larger (M = 47.3, 95% CI = (42.39, 52.22)) than when both target and distractor had the same size (M = 44.2, 95% CI = (39,6, 49.6)), $F(1,58) = 5.22$, $p < .05$, $\eta^2 = .08$. In other words, size estimates are relative, even though the target always had the same size. Crucially, we also found a significant effect of Description on size estimates. If a target was referred to with a redundant size modifier, children perceived it as larger (M = 50.5, 95% CI = (44.99, 56.01)) than when the reference did not include such a modifier (M = 41.43, 95% CI = (36.98, 45.88)), $F(1,58) = 47.03$, $p < .001$, $\eta^2 = .45$. This effect was independent of whether the target was visually larger than the distractor (no significant interaction between Description and Size was found).

Interestingly, we did find a significant interaction between Description and Age, revealing that older children are less sensitive to redundant size modifiers than younger ones, $F(1,58) = 5.02$, $p < .05$, $\eta^2 = .08$. This interaction is illustrated in Figure 5, showing that when the target is minimally referred to ("the football") the size estimates of younger children (M = 41.7, 95% CI = (35.4, 48.0)) were almost the same as those of older ones (M = 41.2, 95% CI = (34.86, 47.45)), while overspecified descriptions ("the large football") caused younger children to make larger estimates (M = 53.74, 95% CI = (45.94, 61.54)) than older ones (M = 47.26, 95% CI = (39.48, 55.06)). When we conducted a separate analysis over the eight additional items which were referred to using non-redundant modifiers, no significant age differences in size estimates were found either, suggesting that it is indeed only redundant size modifiers for which younger children are more sensitive.

No further significant main effects or interactions were found.

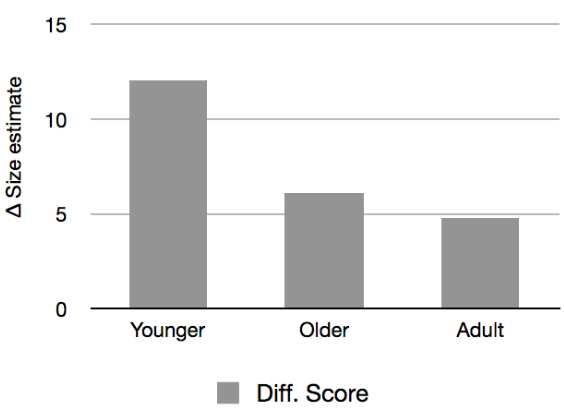

Figure 6: Average difference in size estimate in millimeters for redundant and minimal description ("the large X" minus "the X") as a function of age (comparing younger children, older children and adults).

General discussion

In two experiments we have shown that when target objects are referred to using descriptions containing redundant size modifiers (e.g., "the large red chair" in a situation where two chairs are equally large but different in color), participants subsequently estimate the object to be larger than when objects are referred to using minimally distinguishing descriptions (e.g., "the red chair"). In Experiment 1, we showed this effect with adult language users and with two different size modifiers ("large" and "small"). In Experiment 2, the same effect is shown for children of two different age groups (7- and 10-year olds), but this time with descriptions for object of different types (e.g., a ball and a teddy bear). Interestingly, this effect was found to be independent of whether the target was actually larger than the distractor or not; in both cases, a redundant size modifier had a comparable effect.

Even though the two experiments were slightly different in the way they were conducted (e.g., furniture targets and digital size estimates in Experiment 1 versus children's toys and size estimates on paper in Experiment 2), the essential idea was the same: participants had to process object description which were either overspecified or minimal, and after the objects had disappeared from view, they were asked to estimate the perceived size of the target object that had just been referred to on a scale from 1 to 100. Therefore, it is interesting to plot the difference in size estimates for overspecified and minimal description as a function of age, as is done in Figure 6. Inspection of this figure reveals a clear trend. We already saw in Experiment 2 that 7-year olds were more sensitive to redundant information than 10-year olds, but the pattern for the 10 years olds seems comparable to that of the adults in Experiment 1. A univariate ANOVA confirms this: overall, there is a significant effect of age on the difference in size estimates, $F(2, 125) = 5.91$, $p < .01$, $\eta^2 = .09$, but pairwise comparisons using the Bonferroni method showed that only the younger children differ significantly from the other two age groups. This appears to be consistent with the earlier work (cited in the introduction) showing that children younger than 8 still regularly produce referring expressions that may be underspecified and do not benefit from overspecified descriptions in target identification.

In this study we concentrated on redundant size modifiers, and the question naturally arises whether different kinds of adjectives could have similar effects. Our experiences with color adjectives in Experiment 1 suggest that this may not be the case. In particular, hearing a redundant description of a target as "the large red chair" (when both chairs are red) did not cause participants' to perceive the target as 'redder' than when hearing a minimal description ("the large chair"). Potentially, this could be due to the one-dimensional saturation slider that was used (after all, colors differ along multiple dimensions, also including lightness and hue, and it is not entirely clear which corresponds to, say, 'redness'). Alternatively, it could be due to the fact that color is an absolute property, while size is a relative one (e.g., Rips & Turnbull, 1980). An adjective like "large" implies a comparison (an object is only large compared to another object), which may explain why participants are more likely to modify a size than a color estimate. We leave this as an issue for future research.

Our current results suggest that including redundant modifiers in a referring expression can be used strategically. If a speaker subtly wants to emphasize a property of a target, she can just mention it in a distinguishing description, irrespective of whether it rules out any distractors. In fact, mentioning any redundant property may help in attracting attention to a particular target. Koolen, Krahmer and Swerts (2012), in a study with children from two age groups similar to the ones under study here, found that when children were offered a choice between two identical looking sweets, they opted significantly more often for the one which was referred to in a redundant way ("this red sweet") than for the one that was minimally described ("this sweet"), and even thought the former would taste better than the latter. Interestingly, this effect was found to be stronger for younger than for older children, confirming that as they grow older, and their pragmatic skills increase, children are less likely to be influenced by overspecification.

Conclusion

Wrapping up, we can state that calling a giant both "big" and "friendly" will make him seem larger than merely calling him "friendly", although the size of this effect is presumably inversely proportional to the age of the addressee.

Acknowledgments

Many thanks to Ellen Evers for her help with setting up Experiment 1, and Maud Broekmeulen for her help with conducting Experiment 2. We received financial support from The Netherlands Organization for Scientific Research, (NWO, Vici grant 277-70-007), which is gratefully acknowledged.

References

Ackermann, B., Szymanski, J. & Silver, D. (1990). Children's use of common ground in interpreting ambiguous referential utterances. *Developmental Psychology 26*, 234-245.

Arts, A., Maes, A., Noordman, L. & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43*, 361-374.

Belke, E., & Meyer, A. (2002). Tracking the time course of multidimensional stimulus discrimination: Analysis of viewing patterns and processing times during "same"-"different" decisions. *European Journal of Cognitive Psychology*, *14*, 237-266.

Dahl, R. (1982). *The BFG.* London: Jonathan Cape.

Dale, R. (1989). Cooking up referring expressions. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 68–75.

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science, 18*, 233–263.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*, 1956-1972.

van Deemter, K., Gatt, A., van der Sluis, I. & Power, R. (2012). Generation of referring expressions: assessing the Incremental Algorithm. *Cognitive Science, 36 (5)*, 799-836.

van Deemter, K., Gatt, A., van Gompel, R., & Krahmer, E. (2012). Towards a computational psycholinguistics of reference production. *Topics in Cognitive Science, 4(2)*, 166-183.

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

Engelhardt, P., Bailey, K. & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language, 54*, 554-573.

Engelhardt, P.E., Demiral, S.B., & Ferreira, F. (2011). Over-specified referential expressions impair comprehension: An ERP study. *Brain and Cognition, 77*, 304-314.

Frank, M., & Goodman, N. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, 336, 998.

Grice, H. (1975). Logic and conversation. In: Cole, P. & Morgan, J. L. (Eds.), *Speech Acts*. Academic Press, New York, 41-58.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics, 43,* 3231-3250.

Koolen, R., Goudbeek, M., & Krahmer, E. (2012). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, to appear.

Koolen, R., Krahmer, E., & Swerts, M. (2012). Developmental changes in children's processing of redundant information in definite object descriptions, *submitted*.

Matthews, D., Lieven, E. & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify objects for others: a training study. *Child Development*, *78*, 1744-1759.

Olson, D.R. 1970. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257–273.

Paraboni, I., van Deemter, K. & Masthoff, J. (2007). Making referents easy to identify. *Computational Linguistics, 33*, 229-254.

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

Rips, L. & Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. *Cognition, 8*, 145-174.

Siegal M. & Surian, L. (2004). Conceptual development and conversational understanding. *Trends in Cognitive Sciences, 8*, 534-538.

Sonnenschein, S. (1982). The effects of redundant communication on listeners – when more is less. *Child Development*, *53*, 717-729.