How perceived distractor distance influences reference production: Effects of perceptual grouping in 2D and 3D scenes

Ruud Koolen (r.m.f.koolen@tilburguniversity.edu)
Eugène Houben (eugene@eyetractive.nl)
Jan Huntjens (jan@eyetractive.nl)
Emiel Krahmer (e.j.krahmer@tilburguniversity.edu)
Tilburg Center for Cognition and Communication (TiCC), Tilburg University, The Netherlands

Abstract

This study explored two factors that might have an impact on how participants perceive distance between objects in a visual scene: perceptual grouping and presentation mode (2D versus 3D). More specifically, we examined how these factors affect language production, asking if they cause speakers to include a redundant color attribute in their descriptions of objects. We expected speakers to use more redundant color attributes when distractor objects are perceptually close. Our findings revealed effects of perceptual grouping, with speakers indeed using color more often when all objects in a scene were in the same perceptual group as compared to when this was not the case. An effect of presentation mode (whether scenes were presented in 2D or in 3D) was only partially borne out by the data. Implications of our results for computational models of reference production are discussed.

Keywords: Reference production; overspecification; 2D and 3D scene processing; perceptual grouping; artificial agents.

Introduction

Definite object descriptions (such as "*the red chair"*) are an important part of everyday communication, where speakers often produce them to identify objects in the physical world around them. To serve this identification goal, descriptions have to be unambiguous, and must contain a set of attributes that jointly exclude the *distractor objects* with which the listener might confuse the *target object* that is being referred to. For example, imagine that a speaker wants to describe the object that is pointed at with an arrow in Figure 1.

Figure 1: An example visual scene.

Solving the referential task here requires *content selection*: the speaker must decide on the attributes that she includes in order to distinguish the bowl from any distractor object that is present in the scene (such as the other bowl, the plate and the chairs). This notion of content selection does not only reflect human referential behavior, but is also at the heart of computational models for Referring Expression Generation. Such models, most notably the Incremental Algorithm (Dale & Reiter, 1995), typically seek attributes with which a target object can be distinguished from its surrounding distractors, aiming to collect a set of attributes with which any distractor that is present in the scene is ruled out (Van Deemter, Gatt, Van Gompel, & Krahmer, 2012).

So what would a description of the target object in Figure 1 look like? The target's *type* is probably mentioned because it is necessary for a proper noun phrase (Levelt, 1989). Also *size* is likely to be included, to rule out the large bowl. What else? The speaker may also add *color*, following the general preference to mention this attribute (e.g., Pechmann, 1989), or because the speaker is triggered by the different colors of the objects present in the visual scene (Koolen, Goudbeek & Krahmer, 2013a). In any case, adding color would cause the description to be *overspecified*, since it is not necessary for unique identification: mentioning type and size (*"the small bowl"*) rules out all possible distractors.

If color variation can trigger a speaker to use a *redundant* color attribute, this implies that the distractors in a particular scene largely determine the process of content selection. For the case of Figure 1, it might well be that the speaker would only add color if she were to regard all objects in the scene as relevant distractors (uttering *"the small green bowl"* as a final description). However, there are reasons to believe that speakers tend to ignore certain distractors (Koolen, Krahmer & Swerts, 2013b), and only consider the objects that are into their focus of attention (Beun & Cremers, 1998). This may cause the speaker to leave out color in her description of the target in Figure 1: if she were to restrict her focus space to, say, the two bowls (thereby ignoring the yellow plate), she would probably be less prone to redundantly use color in her description.

What determines whether the yellow plate (or any object in general) is in the speaker's focus of attention? Intuitively, *physical distance* plays a role here: the distant distractor (the plate) might well be ignored, while the closest one (the large bowl) might actually be considered a relevant distractor. In recent empirical research, some evidence has been found for this suggestion (e.g., Clarke, Elsner & Rohde, 2013), though other papers suggest a more nuanced picture (e.g., Koolen et al., 2013b).

In the current paper, we explore two possible factors that may influence *perceived distractor distance* (i.e., perceptual

grouping, and 2D vs. 3D presentation mode), and examine how they influence language production.

Perceptual grouping

The first factor we expect to affect how speakers perceive distance between objects in a scene is *perceptual grouping*. This phenomenon is part of the Gestalt laws of perception (originally introduced by Wertheimer in 1923), and can be defined as people's ability to organize the visual world they perceive in meaningful groups (Palmer, 1992). Among other things, people use this ability to create groups of *objects*, for example when using an expression such as *"the silverware on the counter"*. Thórisson (1994) explains that all kinds of factors can cause people to perceive objects as groups. The most important factors are proximity (where objects that are close together share a group) and similarity (where objects that are similar in shape, color, orientation or function are perceived as a group). Palmer (1992) mentions the principle of common region, which holds that objects that are located together in a *common region of space* are usually perceived as a group (e.g., if they lie within an enclosing contour, such as a table surface).

The question is to what extent perceptual grouping guides speakers in restricting the set of relevant distractor objects in a given scene. Our study provides systematic manipulations of grouping to test this. We hypothesize that objects that are in the same perceptual group as the target are more likely to be in the speakers' focus of attention (in the sense of Beun and Cremers, 1998), and are therefore considered a relevant distractor. Along similar lines, we expect the opposite to be true for objects that are in a different group as compared to the target. Following these expectations, speakers would not consider the yellow plate a relevant distractor in Figure 1, since it is part of a different region of space (the sideboard) than the target (which is placed on the table). Thus, in cases such as these, it is less likely that speakers redundantly use color than when both the target and the distractor are in the same perceptual group.

Presentation mode: 2D vs. 3D scenes

The second factor we expect to affect people's perception of distractor distance relates to how visual scenes are presented to them. In perceiving depth information, people mainly rely on *binocular* depth cues that can only be perceived with two eyes (Loomis, 2001). For the perception of distance between objects, *stereopsis* is an important binocular cue. Stereopsis holds that people view the world from two different angles (one for each eye), which delivers them with two images of a situation. The difference between these two images allows the viewer to perceive distance between objects: if an object is far away, this difference is relatively small, but it is bigger for close objects. Also artificial 3D presentation techniques use two images, thus relying on stereopsis as well.

As far as we are aware, most (if not all) previous work on reference production used flat 2D images (i.e., drawings or realistic photographs) as stimulus material. For such images, viewers depend solely on *monocular* cues (such as relative size, occlusion, and perspective) to perceive distance and depth. Previous work on visual perception has shown that people usually have no difficulty in understanding the three-dimensional nature of 2D images (Saxena, Sun, & Ng, 2008). However, at least for children, it has also been shown that binocular depth and perception is more accurate than monocular depth perception (Granrud, Yonas, & Petterson, 1984), and that 3D scenes are rated higher on naturalness than 2D scenes (Seuntiëns et al., 2005).

The above literature suggests that people are better able to accurately perceive distance between objects in 3D than in 2D visual scenes. Therefore, we hypothesize that the mode of presentation may also affect speakers in determining the set of relevant distractors for a given scene. For example, in Figure 1, the plate might be considered a relevant distractor in 2D, but not in 3D, since speakers might perceive the distance between the target bowl and the plate as bigger in the latter case.

The current study

We performed a reference production experiment, where we presented participants with scenes like the one displayed in Figure 1, and asked them to produce a unique description of a target referent. Crucially, the scenes were set up in such a way that color was never needed to identify the target. This allowed us to take the *proportional use of redundant color attributes* as our dependent variable (following recent work by Koolen et al. (2013b)).

We used two presentation modes to present the stimuli to the participants (2D and 3D), and applied a manipulation of perceptual grouping by systematically placing one distractor (that always had a different color as compared to the target) either in the same region as the target, or in a different one. Third, we replicated a factor that has already been shown to determine speakers' composition of the distractor set, which is related to *distractor type* (Koolen et al., submitted).

We expect, as explained above, that speakers use color more often in the same group condition than in the different group condition. Secondly, we expect speakers to use color more often in 2D than in 3D scenes, because speakers may rely on a bigger distractor set in the former case (due to their poorer estimations of distance).

Experiment

Method

Participants Forty-eight undergraduate students (33 female, mean age: 21.6 years) from Tilburg University took part in the experiment for course credit. All were native speakers of Dutch, the language of the experiment.

Materials The stimulus materials were near-photorealistic visual scenes, modeled and rendered in Maxon's Cinema 4D (a 3D modeling software package[1]). There were 98 trials in total, all following the same basic set-up: participants saw a

[1] See http://www.maxon.net/ for downloads and more information.

Figure 2: Examples of critical trials (in 2D). The left scenes are trials in the same group condition, while the right scenes are trials in the different group condition. The upper scenes are trials in the different type condition, while the lower ones are trials in the same type condition. Note that the trials were presented to the participants on a big television screen.

picture of a living room that contained a dinner table and a sideboard (plus some clutter objects to make the scenes look realistic). The table and the sideboard formed two surfaces on which objects were positioned: one target object and two distractor objects were present in every scene. The target object always occurred at the left side of the table (from the participants' point of view), and had one distractor placed next to it (either left or right). This distractor had the same type and color as the target (meaning that it could only be ruled out by means of its size). Each scene also contained a second distractor – always in a different color as compared to the target – by means of which two principal factors in the design were manipulated (related to perceptual grouping and type). We explain these in more detail below, as well as a third factor (manipulating presentation mode).

Firstly, there was a manipulation of *perceptual grouping*. This factor was manipulated as follows: in half of the trials, the second distractor and the target object were in the *same* group (meaning that they were both positioned on the table), while they were in a *different* group in the other half of the trials (with the target placed on the table, and the distractor on the sideboard). Example scenes for these two conditions can be found in Figure 2. The left scenes represent the *same* group condition: in these scenes, all objects are on the table. The right scenes represent the *different* group condition: the target object (the small bowl) is again on the table, while the second distractor (i.e., the plate in the upper picture, and the yellow bowl in the lower picture) is placed on the sideboard. Crucially, the physical distance between the target and the second distractor was the same in the two conditions.

The second manipulation was related to the *type* of the second distractor in the scene: this could either be different or the same as the target's type. For example, in Figure 2, the second distractor (the plate) has a different type than the target (the bowl) in the upper two trials, while all relevant objects are of the same type in the lower trials. Note also that mentioning a target's type and size was sufficient to distinguish the target in all four scenes, implying that the use of color would always result in overspecification. This applied to all scenes used in the experiment.

The experiment consisted of ninety-eight trials, sixteen of which were critical trials. As said, with regard to the critical trials, all scenes had the same basic set-up, but four different sets of objects were used as target and distractor objects. In Figure 2, trials for one of these sets are depicted (with bowls and a plate). With regard to the other sets, we made sure that they all consisted of food-related objects (such as mugs and cutting boards) that can reasonably be found on a sideboard or a dinner table in a living room. The scenes for these sets of objects were manipulated in a 2 (*perceptual grouping*) x 2 (*type*) design, which resulted in four within conditions as described above: one scene in which the second distractor object shared a group with the target, but not its type; one in which the distractor shared its group and its type with the target object; one in which the distractor neither shared a group, nor its type with the target; and one in which the

distractor did not share its group with the target, but did share its type. The similar first distractor was added to make sure that mentioning type and color was never sufficient to distinguish the target.

Besides the factors perceptual grouping and type, which were both manipulated within participants, we also included one between factor, related to *presentation mode* (2D / 3D). Participants were randomly assigned to either the 2D or the 3D condition. In the 2D condition, the trials were presented to the participants as flat 2D images (i.e. regular photos). As we have explained in the introduction section of this paper, for 2D images, a viewer depends solely on monocular cues to perceive depth information (and distance between objects in particular). In the 3D condition, the trials were presented as 3D images, where speakers could rely on both monocular and binocular depth cues to perceive depth information. The visual scenes in the 2D condition were rendered in the same way as those in the 3D condition, but the image for the left eye was 100% identical to that for the right eye, eliminating depth differences. This means that the 2D and 3D scenes did neither differ in terms of the objects that were visible, nor in the positioning of these objects in the scenes. Moreover, the stimuli as a whole had the same size in the two conditions.

The experiment had eighty-two fillers, all following the setup of the critical trials, with all kinds of objects placed on a table and a sideboard. Again, one of the objects served as the target and was described by the participants, with the crucial difference that the objects in the filler scenes did not differ in terms of their color. In this way, the speakers were discouraged from using color when describing the fillers.

Procedure The experiment took place in an office room at Tilburg University, and participants took part one at a time. The running time for one experiment was approximately 15 minutes. After participants had entered the room, they were randomly assigned to the 2D or 3D condition (there were 24 speakers in both conditions). Thereafter, they were asked to sit down and read an instruction manual. It was explained to the participants that they would be presented with scenes in which one of the objects was marked with an arrow. This target had to be described in such a way that a listener could distinguish it from the other objects that were present in the scene. Once participants were done reading the instructions, they were given the opportunity to ask questions.

The participants (all acting as speakers in the experiment) were seated in front of a large 3D television, while wearing 3D glasses. This was done regardless of the condition they were assigned to, to eliminate differences in the procedure. In the 2D condition, the television displayed flat 2D images of the stimuli. In the 3D condition, the TV used 'active' 3D technology to display the trials: it synchronized with the 3D glasses by means of infrared signals, and used electronic shutters to separate images through the participant's right and left eye. The three-dimensional input was configured as side-by side: both eyes would view an image with a source resolution of 960 by 1080 pixels, presented on an LCD panel with a resolution of 1920 by 1080 pixels. The scenes were presented as still images at 120 Hz, resulting in 60 Hz per eye. In both conditions, participants were shown a short introduction movie (a fragment from the 'Shreck' or 'Ice Age' movies), so that they could get accustomed to the TV and the glasses.

There were two versions of the experiment in terms of trial order: we made one block of trials in a fixed random order (which was presented to half of the participants), and a second block containing the same trials but in reverse order (which was presented to the other half of the participants). The trials were set as slides, and presented using Keynote. No transitions or black screens were used; when a trial was completed, the transition to the next trial was instant. The participants could take as much time as needed to provide a description for every target object, and their descriptions were recorded with a voice recorder. The listener – who was a confederate of the experimenter – sat behind a laptop (out of the speaker's sight), and clicked objects he thought the speaker was referring to. Each time the listener had done this, the next trial appeared. The speaker's instructions told that the listener did not see the stimuli in the same way as the participants, and that the positioning of the objects was different. This eliminated the use of location information as an identifying target attribute, avoiding descriptions such as *"The bowl at the right side of the table"*. The listener never asked clarification questions, to make sure that the speakers produced initial target descriptions.

Design and statistical analysis The experiment had a 2 x 2 x 2 design with two within participants factors[2]: *perceptual grouping* (levels: same, different) and *distractor type* (levels: same, different), and one between participants factor: *presentation mode* (levels: 2D, 3D). The dependent variable was the proportional use of redundant color attributes. As described above, we ensured that participants never needed color to distinguish the target referent from its distractors: mentioning a target's size ruled out the first relevant distractor, while adding the target's type eliminated the second relevant distractor. Thus, if speakers used color anyway, this inevitably resulted in overspecification.

Our statistical procedure consisted of Repeated Measures ANOVAs: one on the participant means ($F1$) and one on the item means ($F2$). To generalize over participants and items simultaneously, we also calculated *MinF'*; we only regarded effects as reliable if $F1$, $F2$, and *MinF'* were all significant. To compensate for departures from normality, we applied a standard arcsin transformation to the proportions before we ran the ANOVAs. For the sake of readability, we report the untransformed proportions in the results section.

[2] Besides the factors mentioned here, the design also contained a replication of one of the factors reported in Koolen, Krahmer, and Swerts (2013b), related to *physical* distractor distance. For this factor, there were trials that either had a close or a distant second distractor object (which were in both cases positioned on the table surface). In line with Koolen et al., there were no differences in the proportional use of color for these two conditions. Due to lack of space, we do not report on these results in the paper.

Results

A total of 768 descriptions were produced in the experiment for the critical trials. These were all fully distinguishing, and speakers mentioned a redundant color attribute in 66,0% of the cases. The order in which the trials were presented to the participants (regular vs. reversed) had no effect on the use of color, and is therefore not further analyzed below.

Results for presentation mode We first examined whether the way in which the trials were presented to the participants (i.e., in 2D or in 3D) had an effect on the redundant use of color. The results show that the presentation mode to some extent affected the use of the redundant attribute color, but this effect was only significant by items ($F1_{(1,46)} = 2.73$, $p = .11$, $\eta_p^2 = .06$; $F2_{(1,12)} = 39.71$, $p < .001$, $\eta_p^2 = .77$; $minF'_{(1,52)} = 2.55$, $p = .11$). This means that the speakers in the 2D condition ($M = .75$, $SE = .07$) included color more often than speakers in the 3D condition ($M = .57$, $SE = .07$), but that we did not find a reliable effect for presentation mode.

Results for perceptual grouping The second factor that we expected to have an effect on the redundant use of color was perceptual grouping. The results indeed showed an effect of grouping on the redundant use of color ($F1_{(1,46)} = 7.81$, $p = .008$, $\eta_p^2 = .15$; $F2_{(1,12)} = 9.02$, $p = .01$, $\eta_p^2 = .43$; $minF'_{(1,41)} = 4.18$, $p < .05$). More specifically, as predicted, we found higher proportions of color use in the *same* group condition ($M = .69$, $SE = .05$) than in the *different* group condition ($M = .62$, $SE = .05$). Overall, this means that our speakers were more likely to include color in scenes where all objects were positioned on the table, as compared to the scenes in which the second distractor was placed on another surface (i.e., the sideboard).

Further inspection of the data suggests that this effect of perceptual grouping was stronger for 3D stimuli rather than 2D stimuli. As visualized in Figure 3, in the case of the 2D stimuli, there was hardly a numerical difference between the *same* group condition ($M = .76$, $SE = .08$) and the *different* group condition ($M = .74$, $SE = .07$), while this difference was bigger for the 3D stimuli (*same* group condition: $M = .63$, $SE = .08$; *different* group condition: $M = .52$, $SE = .07$). However, this interaction between perceptual grouping and presentation mode only reached significance by participants ($F1_{(1,46)} = 4.61$, $p = .04$, $\eta_p^2 = .09$; $F2_{(1,12)} = 2.97$, $p = .11$, $\eta_p^2 = .20$; $minF'_{(1,29)} = 1.80$, $p = .19$). Therefore, this interaction was not statistically reliable.

Results for distractor type Thirdly, we aimed to replicate the effect of *type* (reported on by Koolen et al. (submitted)) expecting the type of the second distractor to have an effect on redundant color use. Distractor type indeed had an effect on the redundant use of color ($F1_{(1,46)} = 6.88$, $p = .01$, $\eta_p^2 = .13$; $F2_{(1,12)} = 9.09$, $p = .01$, $\eta_p^2 = .43$; $minF'_{(1,44)} = 3.91$, $p = .05$). This means that speakers more often used color when the distractor's type was the same as the target's type ($M = .69$, $SE = .05$) as compared to when its type was different ($M = .63$, $SE = .06$).

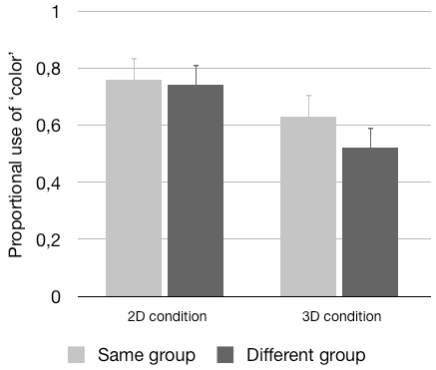

Figure 3: The proportional use of color (plus standard deviations) for the 2D and 3D conditions as a function of the same group and different group stimuli.

Discussion

In the current paper, we studied how the *perceived distance* between objects in a scene affects speakers' production of definite object descriptions, and, in particular, to what extent it causes them to include redundant color attributes in such descriptions. Firstly, we replicated the effect of *distractor type*, reported earlier by Koolen et al. (submitted): we found speakers to use color more often when a target object and a differently colored distractor were of the same type (e.g., two bowls) as compared to when they had different types. These findings suggest that an object is more likely to be considered a relevant distractor if it shares its type with the target (as compared to when this is not the case).

Our findings did not reveal reliable effects of *presentation mode* (2D vs. 3D) on redundant color use. We hypothesized that it is more difficult for people to accurately perceive the distance between a target object and a given distractor in 2D scenes rather than in a 3D version of the same scenes, since in a 3D presentation mode, speakers can use both monocular and binocular cues for depth perception (Loomis, 2001). We indeed found a numerical difference (in terms of redundant color use) between the conditions in the expected direction, but this difference only reached significance by items. One explanation for this could be related to the way in which we manipulated distance between objects in the scenes: this was done horizontally, on the X-axis. It may be that the effect of presentation mode is stronger when distance is manipulated along the depth (Z) axis, or along the X-axis and the Z-axis at the same time. Arguably, in the latter cases, the difference between actual and perceived distance may be interpreted as bigger in 3D than in 2D. In future research, we plan to study if this is indeed the case.

With regard to our manipulation of *perceptual grouping*, we were able to confirm our expectations. We hypothesized that objects that are in the same region of space as the target are more likely to be considered as a relevant distractor than objects in a different region of space (in the sense of Palmer, 1992). To test this, we systematically placed one distractor (the one with the different color) either in the same region as the target, or in a different one (keeping the actual distance between the objects the same). Participants used color more

often in the *same* group condition than in the *different* group condition, which suggests that the differently colored object was more likely to be in a speakers' focus of attention (Beun & Cremers, 1998) in the former case. Along the lines of Palmer (1992), our findings imply that speakers indeed tend to perceive objects around them in groups, and that this tendency guides them in determining the distractor set when describing objects in a scene. In future research, we plan to validate this suggestion by collecting eye-tracking data, and to extend the results reported on here with manipulations of grouping other than region of space, such as proximity and similarity (see also Casasanto, 2008). Furthermore, also the interaction between grouping and presentation mode would be worth exploring in future research: although it seemed to be the case that the effect of grouping was practically absent in 2D and strong in 3D, this interaction was only significant in $F1$, but not in $F2$ and $MinF'$ (and therefore not reliable).

The finding that people rely on perceptual grouping when determining the set of distractors for a scene has interesting implications for current computational models in the field of Referring Expression Generation (REG), most notably Dale and Reiter's (1995) Incremental Algorithm (IA). As noted in the introduction, such models are artificial agents that aim to generate distinguishing descriptions of objects, and compute a set of attributes that rules out all distractors in a given scene. However, for their IA, Dale and Reiter (1995, p. 236) define the distractor set as "the set of entities that the hearer is currently assumed to be attending to". This means that the IA normally includes any object that is present in a scene in the distractor set, following many other algorithms in the field. However, while Krahmer and Theune (2002) show that the distractor set that REG algorithms use may change during a discourse, our findings for perceptual grouping suggest that the region in which objects occur should be taken into account as well: objects that do not share their region with the target should not always be considered.

Conclusion

This paper explored the impact of perceptual grouping and presentation mode (2D versus 3D) on how people perceive distance between objects in a visual scene when referring to objects. The results showed an effect of perceptual grouping on the redundant use of color, implying that objects that are in the same region of space as the target are more likely to be considered a relevant distractor than objects that are in a different region. Our manipulation of presentation mode did not reveal reliable effects on redundant color use.

Acknowledgments

We thank Fons Maes, Hans Westerbeek, Martijn Goudbeek, and Jorrig Vogels for their comments on an earlier draft.

References

Beun, R.J., & Cremers, A. (1998). Object reference in a shared domain of conversations. *Pragmatics & Cognition, 6* (1/2), 121-152.

Casasanto, D. (2008). Similarity and proximity: when does close in space mean close in mind? *Memory & Cognition, 36* (6), 1047-1056.

Clarke, A., Elsner, M., & Rohde, H. (2013). Where's Wally: the influence of visual salience on referring expression generation. *Frontiers in Psychology, 4*, article 329.

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

Granrud, C., Yonas, A., & Pettersen, L. (1984). A comparison of monocular and binocular depth perception in 5- and 7-year-old infants. *Journal of Experimental Child Psychology, 38* (1), 19-32.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013a). The effect of scene variation on the redundant use of color. *Cognitive Science, 37* (2), 395-411.

Koolen, R., Krahmer, E., & Swerts, M. (2013b). The impact of bottom-up and top-down saliency cues on reference production. In *Proceedings of the 35th annual meeting of the Cognitive Science Society (CogSci)*, 817-822. Berlin, Germany.

Koolen, R., Krahmer, E., & Swerts, M. (submitted). How distractor objects trigger referential overspecification: testing the effects of visual clutter and distance.

Krahmer, E., & Theune, M. (2002). Efficient context-sensitive generation of referring expressions. In: K. van Deemter, & R. Kibble (Eds.). *Information sharing: givenness and newness in language processing* (pp. 223-264). CSLI Publications, Stanford.

Levelt, W. (1989). Speaking: from intention to articulation. MIT Press, Cambridge/London.

Loomis, J. (2001). Looking down is looking up. *Nature, 414* (6860), 155-156.

Palmer, S. (1992). Common region: a new principle of perceptual grouping. *Cognitive Psychology, 24* (3), 436-447.

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

Saxena, A., Sun, M., & Ng, A. (2008). Make3D: depth perception from a single still image. In *Proceedings of the 23rd AAAI conference on artificial intelligence*, 1571-1576. Chicago, Illinois, USA.

Seuntiëns, P., Heynderickx, I., IJsselstein, W., Avoort, P., Berentsen, J., Dalm, I., …, Oosting, W. (2005). Viewing experience and naturalness of 3D images. In *Optics East*, Boston, Massachusetts, USA.

Thórisson, K. (1994). Simulated perceptual grouping: an application to human-computer interaction. *Proceedings of the 16th annual conference of the Cognitive Science Society (CogSci)*, 876-881. Atlanta, Georgia, USA.

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

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, 4, 301-350.