Alignment of (dis)preferred properties during the production of referring expressions

Martijn Goudbeek (m.b.goudbeek@uvt.nl)
Department of Communication and Cognition, University of Tilburg, Warandelaan 2
5000 LE, Tilburg, The Netherlands

Emiel Krahmer (e.j.krahmer@uvt.nl)
Department of Communication and Cognition, University of Tilburg, Warandelaan 2
5000 LE, Tilburg, The Netherlands

Marc Swerts (m.g.j.swerts@uvt.nl)
Department of Communication and Cognition, University of Tilburg, Warandelaan 2
5000 LE, Tilburg, The Netherlands

Abstract

It is generally assumed that speakers prefer certain properties over others when referring to targets, and various algorithms for the generation of referring expressions model this assumption. These preferences are often claimed to be fixed for a given domain. We argue that this is not the case, and that speakers' preferences are at least in part determined by the preferences of other speakers. In an experiment in which participants could describe people or furniture pieces based on preferred and dispreferred properties, we found that speakers frequently aligned with respect to the previously heard (preferred or dispreferred) property, even though they could always refer with the preferred property. Speakers aligned more with the dispreferred property in the more clearly structured furniture domain.

Keywords: Alignment; Referring Expressions; Property selection; Incremental Algorithm

Introduction

Speakers often produce referring expressions such as *The man with the beard* or *The red chair*, and this production process has been studied extensively both from an experimental and from a computational perspective. Within the computational perspective, many researchers have addressed the question how to automatically determine which properties to include in a "distinguishing description" (i.e., a description that uniquely characterizes a target by distinguishing it from a set of distractor objects). Dale and Reiter's (1995) Incremental Algorithm is often considered the algorithm of choice for this problem, due to its algorithmic simplicity and empirical groundedness, since it incorporates psycholinguistic princi-

ples such as those uncovered by Pechmann (1989) and pragmatic principles as put forward by Grice (1975). A basic feature of the Incremental Algorithm is the prediction that speakers enter distinguishing properties into their descriptions based on a preference list. The order of properties in this preference list is assumed to be fixed for a given domain, so that speakers are predicted to invariably prefer using certain properties over others. As a consequence, they will only use dispreferred properties when the preferred properties are not sufficient to uniquely characterize the target. For example, when given the choice to distinguish a blue front-facing chair from a red sideways-facing chair, speakers are likely to use the preferred property color and not the dispreferred property orientation (Gatt, van der Sluis, & van Deemter, 2007). Only when both chairs would have the same color would speakers start using orientation in their descriptions.

We argue that this strategy may work for descriptions in monologues, but is arguably not a good model for the production of referring expressions in interaction. This claim is based on a common observation from experimental work on reference, where it has been shown repeatedly that speakers tend to align with each other with respect to the expressions they may use to refer to objects (Brennan & Clark, 1996). For example, if a speaker refers to a couch using the word *sofa* instead of the more common *couch*, the addressee is more likely to use *sofa* instead of *couch* as well later on in the dialogue (Branigan, Picker-

ing, Pearson, & McLean, in press). This kind of alignment ("entrainment") is well-established for lexical choice, and this lexical phenomenon is perfectly compatible with the Incremental Algorithm. After all: both *sofa* and *couch* are alternative realizations of a single underlying property (i.e., being an article of furniture for sitting), and the Incremental Algorithm only operates on the level of property selection. However, the question we address in this paper is whether speakers align with each other on the level of which properties they select to refer to objects? When speaker A uses a dispreferred property to refer to a target, will speaker B use that same dispreferred property, even when a more preferred property is available? We hypothesize that, contrary to the predictions of the Incremental Algorithm, the preference list as observed in human speakers is not fixed for a given domain, but varies with the properties that speakers have been exposed to in the preceding dialogue. Here we present experimental evidence for this hypothesis, and discuss ways of how referring expression generationj algorithms could be extended to incorporate this finding, namely by introducing plasticity in the properties' preference order.

To investigate the role of alignment processes during the production of referring expressions, we constructed an experiment in which participants alternatively had to identify a target based on a description they heard and had to refer to a different target themselves so that another listener could identify it. In the critical trials, the descriptions the participants heard were either with preferred properties (e.g., *The man with the glasses, the blue desk* or with dispreferred properties (e.g., *The man with the tie, the desk seen from the side*). We investigated alignment to both preferred and dispreferred properties to test the predictions of the Incremental Algorithm. Recall that this algorithm posits a preference list: properties that rank high on the list are selected first and properties low on the list (i.e., dispreferred properties) will be selected only if they are necessary to uniquely identify an object.

The dependent factor is the extent to which the participants are sensitive to whether a preferred or dispreferred property is being used in the descriptions they hear.

Method

Participants

Twenty-six students (two males, mean age = 20.96, SD = 2.23) from the Radboud University of Nijmegen participated as part of a course requirement. All were native speakers of Dutch and had no history of hearing or speech problems.

Materials

The target pictures were taken from the TUNA corpus (Gatt et al., 2007) that has been extensively used in the study of referential expression. This corpus consists of two sub-domains of pictures: a domain containing pictures of people (portraits of mathematicians who could, for instance, be described as *The bald man with the glasses*) and a domain containing pictures of furniture items[1] in different colors depicted from different orientations (e.g., *The red desk facing left*).[2]

For the current experiment, it is important that participants in previous research (Gatt et al., 2007; Koolen, Gatt, Goudbeek, & Krahmer, 2009) did have a preference for certain properties when referring to targets in these domains (e.g., color in the furniture sub-domain, glasses in the people sub-domain) and a dislike for certain other properties (e.g., orientation of a furniture piece and the wearing of a tie). Given the choice, participants prefer to say *The green sofa* when they could also have described the picture with *The sofa seen from the side* in the furniture set.

Procedure

Every trial consisted of four tasks: a prime, two fillers and the experimental description. Figure 1 depicts these four tasks. First, participants got the prime: they had to listen to a preferred or dispreferred description and indicate (by button press) which of the three presented pictures was referred to. Second, they were asked to verbally refer to a filler picture (a picture from the other sub-domain) by distinguishing it from two distractors. The participant produces, say, *The man with the white hair*

---

[1]The picture of furniture items were taken from the Object Databank, developed by Michael Tarr at Brown University and freely distributed. URL: http://titan.cog.brown.edu:8080/TarrLab/stimuli/objects/

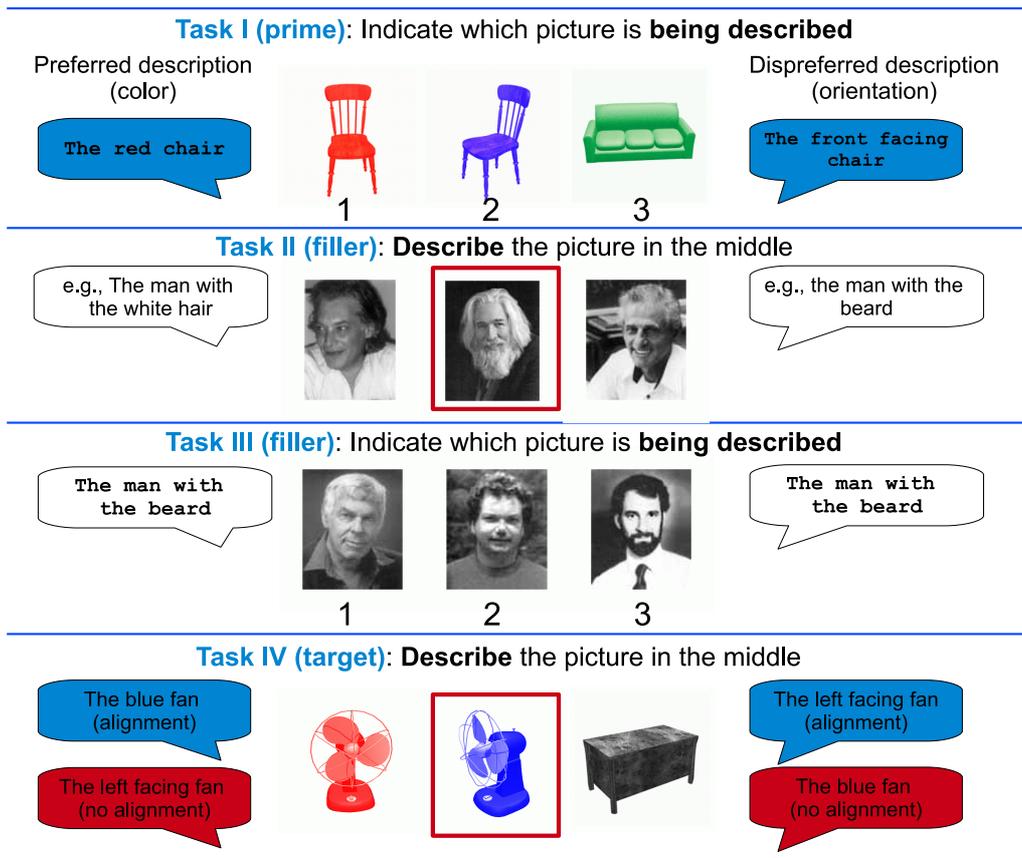[2]Here and elsewhere we give English versions of Dutch originals.

Figure 1: The four stages of a trial in the experimental procedure. Time proceeds from top to bottom. The left side of the figure depicts a trial where the participant was exposed to a preferred property in Task I, the right side depicts a trial where the participant was primed with a dispreferred property. Utterances that are **heard** by the participant are in Courier font, utterances that are **spoken** by the participant are in Arial.

Third, they had to listen to a description of a filler picture and indicate which picture was referred to. Finally, they had to describe the target object, as they did in task two. Crucially, the target object could always be distinguished from the two distractors in two ways: using a preferred (e.g., *The blue fan*) or using a dispreferred property (e.g., *The left facing fan*). Note that the experimental descriptions in the furniture domain are not literal copies of the primes, rather they use the same attributes but with potentially different values. Note also that there were always two filler tasks between the prime and the experimental description to prevent a too direct connection between the prime and the target. The items in these filler tasks always came from the other domain: when the critical trials were about furniture, the filler trials depicted people and vice versa. In addition, half of all the trials were filler trials that did not contain critical descriptions but descriptions without modifying properties (e.g., *The desk, the chair*) or description with filler descriptions (e.g., *The man with the beard*). These items occured solely in the filler trials.

The distance between prime and target and the inclusion of filler trials were intended to prevent participants from guessing the purpose of the experiment. Additionally, the participants were

given a filler task to perform during the experiment. This task consisted in memorizing all the portraits of the bearded men they encountered during the experiment. After the experiment, they had to indicate which faces they remembered on a sheet of paper. In the debriefing following the experiment none of the participants indicated that they had been aware of the researcher's purpose during the experiment.

In total there were 160 trials: 80 trials in each blocks (40 filler trials, 20 critial furniture trials, 20 critical people trials).

Design and analysis

The experiment consisted of two blocks: one block consisting of the preferred furniture descriptions and the dispreferred people descriptions and one block with the preferred people descriptions and the dispreferred furniture descriptions. Within each block, trials from the furniture domain, trials from the people domain and filler trials were randomized. All participants were presented with both blocks: half of the participants started with a randomized block of the preferred furniture and dispreferred people descriptions and the other half started with a randomized block the dispreferred furniture and preferred people descriptions. Thus, the experiment has a two by two design: two factors (domain and stimulustype) with two levels (people and furniture; preferred and dispreferred).

We use two dependent measures to quantify the behavior of our participants: the proportion of alignment and the speech onset time of a trial. Within a trial, there is alignment if the participant uses the same attribute in the referring expression that was encountered in the previous trial. For example, in the furniture domain, when a participant hears a description with orientation (e.g., *The front facing desk*), any description that mentions orientation in the experimental trial (e.g., *The chair facing left*) counts as alignment. Using both the preferred and the dispreferred property in a description (e.g., *The red chair facing left*) is also counted as alignment, although it can also be constructed as overspecification (Engelhardt, Bailey, & Ferreira, 2006; Arnold, 2008) since leaving out one attribute would still result in a distinguishing description. This possible confounding of overspecification and alignment is addressed in the Re-

sults section. In addition to mean percentage of alignment, we measured the time it took for participants to start producing their description, the speech onset time. The idea behind this measure is that the longer speakers take to start speaking the more difficult (more mentally taxing, more dispreferred) the description is. The speech onset time could thus shed light on the processes underlying the behavior in our experiment.

Results

We performed two separate analyses of variance with domain (furniture versus people) and stimulustype (preferred versus dispreferred) as factors and mean proportion alignment and speech onset as dependent variable.

Proportion alignment

The left column of Table 1 shows the proportion of alignment for each domain and stimulustype. As a first test of our hypothesis that alignment processes play an important role in property selection for referring expressions, we investigated participants' use of the dispreferred property.

Table 1: Mean alignment (Align) and overspecification (Over) with their standard deviations per domain (furniture and people) and stimulus type (Type; preferred and dispreferred).

Domain	Type	Align (SD)	Over (SD)
Furniture	Pre	0.89 (0.32)	0.14 (0.34)
	Dis	0.60 (0.49)	0.11 (0.31)
People	Pre	0.97 (0.16)	0.15 (0.36)
	Dis	0.25 (0.43)	0.13 (0.33)

In the preferred stimulustypes, the hypothesized effects of alignment and the preference order of the Incremental Algorithm work in the same direction. In the dispreferred stimulustypes however, the Incremental Algorithm predicts that the preferred property is selected and the dispreferred property will be ignored. This is not what we observe. Instead, there is substantive use of the dispreferred property in stimulustypes where our participants were primed with the dispreferred property. To quantify these observations, we performed a one sample t-test for the dispreferred

stimulustypes with zero as the reference value for the prediction of the Incremental Algorithm. In both domains the use of the dispreferred property significantly exceeds the predictions set by the Incremental Algorithm ($t_{furniture\,dispreferred}$ [25] = 6.86, p < 0.00, $t_{people\,dispreferred}$ [25] = 4.81 p < 0.00).

The right column of Table 1 shows the amount of overspecification for each stimulustype. While there is a considerable amount of overspecification in all stimulustypes and domains, this does seem to be independent of the domain or stimulustype. We investigated this with a two-way repeated measures anova with overspecification as dependent variable and domain and stimulustype as independent variables. This analysis showed that there is no difference in overspecification between the people domain and the furniture domain (F [1,25] = 0.095, *n.s.*) nor between the dispreferred and the preferred stimulustype (F [1,25] = 2.019, *n.s.*). Alignment and overspecification thus seem to be two independent processes. When our listeners aligned with the previously encountered property, they did not necessarily add this property to their descriptions (resulting in overspecification), but used only this property in their referential expressions.
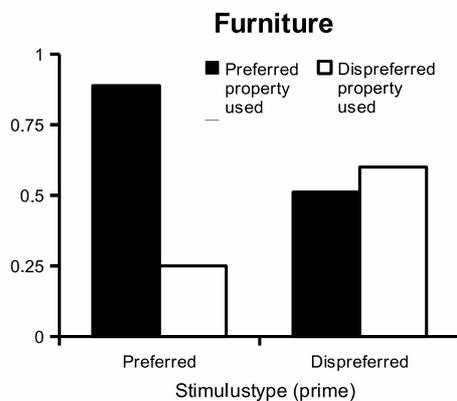

Figure 2: The use of the preferred and dispreferred property for the preferred and dispreferred stimulustypes in the furniture domain.

The use of the preferred and dispreferred property with the different stimulustypes (primes) in the furniture and people domain is shown in Figure
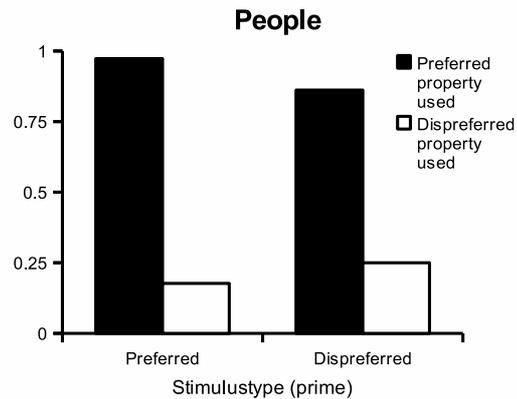

Figure 3: The use of the preferred and dispreferred property for the preferred and dispreferred stimulustypes in the people domain.

2 and Figure 3 respectively. In both domains, the preferred property is used more than the dispreferred property with the preferred stimulustypes. There is an increase in the use of the dispreferred property with the dispreferred stimulustypes, although this increase is markedly smaller in the people domain. These effect were tested statistically in a repeated measures anova with amount of alignment as dependent variable and domain and stimulustype as dependent variables. This analysis showed that there was significantly more alignment in the furniture domain than in the people domain (F [1, 25] = 22.96, p < 0.00) and that our participants aligned more with the preferred property than with the dispreferred property. Additionally, there was a significant interaction between domain and stimulustype (F [1, 25] = 15.00, p < 0.00) indicating that there was more alignment with the dispreferred property in the furniture domain.

Speech onset

In addition to the proportion alignment, we also measured the time it took the participants to start speaking their utterance. This speech onset time can be considered a rough measure of the duration of the utterance preparation phase. Figure 4 shows the speech onset time for the two stimulustypes and the two domains. The comparison of the left and right bars show that the participants were slower in the furniture domain while the compar-

ison of the black and white bars show that participants were slower when they were primed with a dispreferred property. A repeated measures analysis of variance confirmed the slower speech onset in the furniture domain (F [1, 25] = 4.98, p < 0.04) and showed the difference between the preferred and dispreferred stimulustypes to be marginally significant (F [1, 25] = 3.38, p < 0.08). There was no significant interaction between domain and stimulustype (F [1, 25]= 1.24, p < 0.28). A possible explanation for these effects lies in the realization of the dispreferred property in the furniture domain (e.g., *The chair seen from the side*) which is a very infrequent way of phrasing in Dutch. Furthermore, both the preferred and dispreferred description in the people domain start with the same phrase (*The man with the glasses* and *The man with the tie*), which explains the faster speech onset in the people domain.
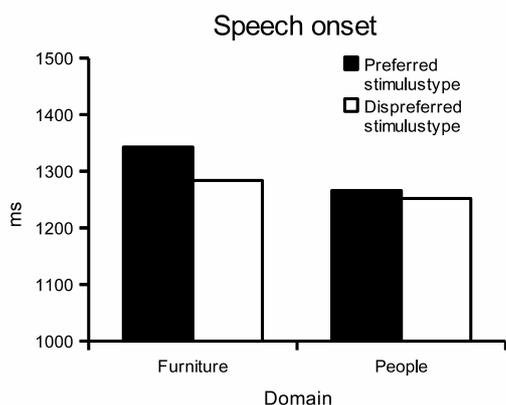

Figure 4: The speech onset time in milliseconds for preferred and dispreferred stimulustypes in the people and furniture domain.

In addition to the effects of domain and stimulustype, we also explored the effect of alignment on speech onset. Since alignment is considered to facilitate the production of language, we hypothesized that trials on which participants aligned with the encountered description would have faster speech onset times, even if this description used a dispreferred property. Figure 5and Figure 6 display the effect of alignment on speech onset time per domain and stimulus type.

We conducted a 2 x 2 x 2 repeated measures anova with speech onset as dependent variable and alignment (absent or present), domain and stimulustype as independent variables.[3]
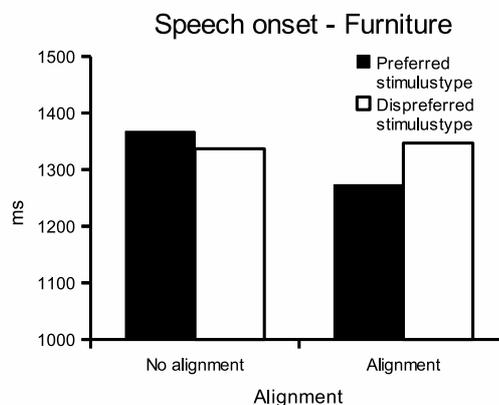

Figure 5: The speech onset time for preferred and dispreferred aligned and non aligned stimuli in the furniture domain.
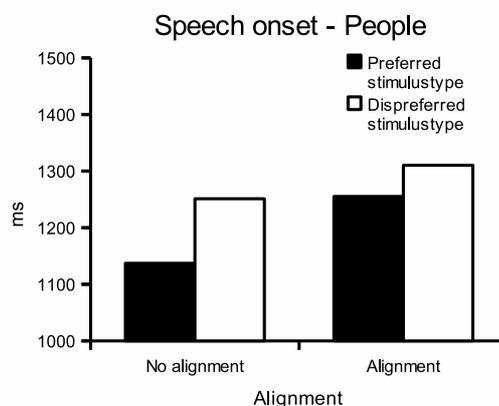

Figure 6: The speech onset time for preferred and dispreferred aligned and non aligned stimuli in the people domain.

---

[3]Because our participants tended to be fairly consistent in whether they aligned with a previously encountered utterance or not, a considerable number of cells in the design are empty. These missing values were replaced by the mean of their series. While this procedure made the analysis feasible, the results need to be cautiously interpreted with this manipulation in mind.

The results of the analysis showed that, contrary to our hypothesis, speakers are often slower in trials where they align with the previously mentioned property. This slower speech onset in alignment trials is due to the consistently large differences in speech onset time in the people domain between trials with and without alignment. Aligning with the dispreferred phrase *The man with the tie* does not benefit the speaker, nor does aligning with the preferred property (*The man with the glasses*). In the furniture domain aligning with the preferred description does result in faster speech onset times, although this could be due to this property (i.e., *color*) being easier to process regardless of alignment benefits. In total, these effects lead to a small but significant effect of alignment on speech onset (F [1, 25] = 4.11, p < 0.05) indicating slower speech onset times when speakers aligned with the encountered description.

Discussion

The results show that that participants frequently align on both dispreferred and preferred properties when producing their referring expressions. This shows that speakers' preferred and dispreferred properties are not fixed for a particular domain, but also depend on the properties they have been exposed to before. Interestingly, participants aligned more with dispreferred properties in the furniture domain than in the people domain. There are several possible explanations for the difference between the two domains. First, it could be due to the greater variability in possible descriptive properties in the people domain compared to the minimally varying furniture domain. The furniture domain systematically varied the properties *type*, *color* and *orientation* whereas the people domain contained faces that can be described in many different ways. We balanced the faces with respect to the relevant properties (*glasses* and *ties*) and with respect to two frequently used properties *(having a beard* and *being bald*), but there are many other properties our participants could have paid attention to. This greater variability in possibly entertained properties might have hampered alignment. In addition, the exact position of a property on the preference list is unknown; while ties and orientation are both dispreferred properties, ties might be more dispreferred than orientation and

thus less prone to alignment effects. Related to this is the possible difference in salience between the two dispreferred properties: differences in the orientation of the furniture items were extremely clear, whereas the ties were sometimes less visually salient.

The results from the speech onset analysis show expected effects of domain and stimulustype. It takes participants less time to start speaking in the people domain, probably because of the lesser variation in the start of the utterances in this domain (they all started with *The man with the...*). It also takes them longer to utter a description with a dispreferred property, especially in the furniture domain. The effects of alignment on speech onset both on domain and stimulustype. In the people domain, speakers are slower when they align, in the furniture domain. speakers are somewhat faster when they align, primarily when they use the preferred property.

In conclusion, the results clearly show alignment effects in the selection of distinguishing properties for referential expressions. Alignment effects depend both on the domain used and whether speakers need to align with a preferred or dispreferred property. Speakers do align with dispreferred properties, but more in the furniture domain than in the people domain.

Algorithms that generate referring expressions, such as the Incremental Algorithm (Dale & Reiter, 1995) and the Graph algorithm (Krahmer, Erk, & Verleg, 2003) can be made more psychologically realistic by implementing dynamic instead of fixed preference lists. The order of the preference list can first be determined by preference (e.g.,word frequency) and subsequently be adjusted based on contextual input.

This experiment dealt with the planning (selection) of distinguishing properties for referring expressions. Future work will deal with the role of alignment in the realization of referring expressions and with the relation between alignment and overspecification in referring expressions.

To investigate the relation between alignment and realization, participants will be confronted with preferred and dispreferred realizations. For example, in the people domain the preferred description (in Dutch) of a man with glasses and a beard could be *The glassed and bearded man,*

while the dispreferred description would be or *the bearded and glassed man*.[4] Similarly, in the furniture corpus the preferred description of a big red chair would be *The big red chair* while the dispreferred description would be *The red big chair*. If alignment processes also influence the realization of referring expressions, speakers will tend to realize their referring expressions in the same way as the primes they heard.

The relation between overspecification and alignment will be investigated by priming participants with overspecified referring expressions (e.g., *The man with the tie and the glasses* or *the red desk seen from the left*) when only one property would be sufficient to uniquely identify the target. We hypothesize that, compared to our current results, participants will align more when the both the preferred and dispreferred property are primed.

Acknowledgments

References

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

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (in press). Linguistic alignment between people and computers. *Journal of Pragmatics*.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482-1493.

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233-263.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, *54*(4), 554–573.

Gatt, A., van der Sluis, I., & van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the 11th european workshop on natural language generation.*

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2009). Need I say more? on factors causing referential overspecification. In *Proceedings of the pre-cogsci 2009 workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.* This volume.

Krahmer, E., Erk, S. van, & Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, *29*(1), 53–72.

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

---

[4]While ungrammatical in English, these different property orderings are possible in Dutch.