Referring under load

Goudbeek, M.B.; Krahmer, E.J.

Published in:
Proceedings of the Workshop on the Production of Referring Expressions (PRE-CogSci 2011)

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Referring under load: Disentangling preference-based and alignment-based content selection processes in referring expression generation

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

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

Abstract

In dialogue, speakers arguably have two ways in which to select attributes for inclusion in their referring expressions; they can rely on a predefined preference order until, or they can resort to attributes used earlier in the interaction. In this paper we study the interplay between these two strategies using a dual task paradigm, where participants are asked to interactively refer to object, while simultaneously performing a demanding secondary verbal task. The results showed that when speakers were under load, they tended to rely less on attributes that were primed in the preceding interaction. The results are discussed as evidence for a dual route model of referring expression, such as that proposed by Gatt, Goudbeek, & Krahmer (2011).

Keywords: Referring expressions, alignment, preferences, attributes, dual task, content selection, overspecification

Introduction

The production of referring expressions can be seen as a problem of choice. Imagine that you want to refer to a chair, and that you can do this in at least two different ways, as, say, *the blue chair* of *the chair seen from the front*. Assume moreover that both these descriptions are distinguishing (that is: both uniquely characterize the target chair by distinguishing it from a set of distractor objects). Which would you chose? Many people would probably opt for the description using color. Indeed, there is substantial empirical evidence that color is often a highly preferred attribute, and more in general, that, given some domain, speakers have preferences for some attributes over others (see, for example, Pechmann, 1989). But what if you are interacting with a person who happens to prefer the attribute orientation over color when discussing furniture items, for example, referring to a desk as *facing left* and a couch as *seen from the back*. Would you then still stick to your own preference for color, or would you align with your dialogue partner and start using orientation as well? The question how speakers perform the balancing act of these two different processes (relying on preferences or aligning with descriptions earlier in the dialogue) is the topic of this paper.

Until recently, most Referring Expression Generation (REG) algorithms tended to ignore dialogue, and work solely on the assumption that some attributes are preferred over others (Krahmer & van Deemter, in press). The Incremental Algorithm (Dale & Reiter, 1995), for example, assumes the existence of a fixed, domain dependent list of preferred attributes, and when producing a referring expression, the algorithm iterates through this list, adding an attribute (e.g., color) to the description under construction if its value (e.g., blue) rules out any of the distractors (because they have a different color).

Recent work by Goudbeek and Krahmer (2010, in press) has challenged the assumption that speakers operate on a fixed, domain dependent preference order when references are produced in a dialogue. In their study, participants alternatingly identify furniture items or persons based on a spoken referring expression (the prime) and refer to a similar item themselves later on in the dialogue (the target). In one experiment the primed descriptions could use preferred (color in the furniture domain) or dispreferred (orientation) attributes, while in another experiment primed descriptions used *both*, resulting in overspecified descriptions such as *the blue chair seen from the*

front. Goudbeek and Krahmer found that participants were much more likely to use dispreferred or overspecified descriptions themselves when these were primed earlier in the interaction. They interpret this as a form of adaptation or alignment at the level of attributes, in line with psycholinguistic studies such as Brennan and Clark (Brennan & Clark, 1996), who argue that interlocutors in dialogue form "conceptual pacts" on how to refer to objects, or Pickering and Garrod (Pickering & Garrod, 2004), who argue that interlocutors automatically align their representations as a result of priming.

Interestingly, most existing REG algorithms (including the Incremental Algorithm) fail to account for these findings, since they predict that the preferred attribute (color) would always be used and the dispreferred attribute (orientation) never, also not redundantly together with color. Therefore, a new computational model was proposed and evaluated for the production of referring expressions (Gatt, Goudbeek, & Krahmer, 2011), combining a preference-based REG model with an alignment based model working in parallel, with both contributing to a limited-capacity working memory (see Figure 1 for a graphical rendition of the model). This model distinguishes between dynamic effects arising during an interaction, and more stable effects such as property preferences, which are likely to be related to properties of the domain of reference and the human perceptual system. These dynamic effects that occur in dialogue have been addressed in computational models of alignment in speech planning (see, for example, Reitter, Moore, & Keller, 2006; and Buschmeier & Bergmann for an implementation of an alignment capable speech planner). . However, these studies have not addressed dynamic effects in referring expression generating in the context of preferences.

The use of a limited-capacity working memory buffer predicts that occupying the buffer should directly affect the production referring expressions. The current paper tests this prediction using a dual-task paradigm, in which participants carry out a memory task while simultaneously performing the reference task. If we find that limiting working memory impacts one process in the model by Gatt et al. (2011) more than the other,

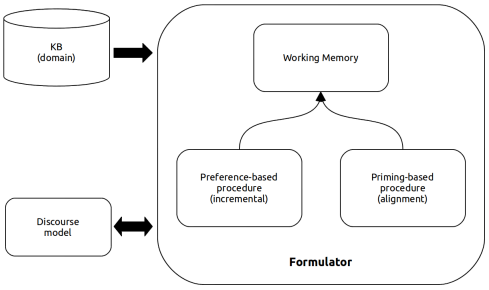

Figure 1: The parallel model with two procedures contributing to the generation of referring expressions: a preference based procedure and a priming based procedure.

this would be compelling evidence for the existence of two separate processes running in tandem. Interestingly, even though reference in interactions has been studied extensively by psycholinguists, the psycholinguistic literature is equivocal regarding what would happen when speakers refer under load. On the one hand, if we assume that processing the utterances of another speaker requires cognitive resources (in line with, for instance, Horton & Keysar, 1996; Keysar, Barr, Balin, & Paek, 1998), then we could expect that participants will tend to rely less on alignment and more on their own preferences. If, however, we assume that attending to the utterances of the other speaker is cheap and automatic (e.g., Pickering & Garrod, 2004), then we would expect that having to perform a dual task has little or no impact on the references that speakers produce.

Method

Participants

In the content selection experiment 26 participants took part (five males, twenty-one females, mean age = 20.66, SD = 2.13) and in the overspecification experiment 28 participants took part (eight males, twenty females, mean age = 20.1, SD = 1.89). All participants were students from Tilburg University and participated in exchange for partial course credit.

Materials

The pictures that had to be referred to were taken from the TUNA corpus (van Deemter, Gatt, van

der Sluis, & Power, in press) that has been extensively used in the study of referential expression. This corpus consists of two sub-domains: a domain containing pictures of people (portraits of mathematicians who could, for instance, be referred to as *The bald man with the glasses*) and a domain containing pictures of furniture items[1] in different colors depicted from different orientations who could be referred to as *The red desk facing left*.[2] In the current experiments, the critical stimuli are all taken from the furniture domain.

In previous research (van Deemter et al., in press; Koolen, Gatt, Goudbeek, & Krahmer, 2009) participants have been shown to have a preference for certain properties when referring to targets in these domains. When referring to furniture items participants preferred color over orientation (e.g., when given the choice, participants prefer to say *The green sofa* when they could also have referred to the picture with *The sofa seen from the side*).

Procedure

For the purposes of the current experiment, we combined the interactive alignment paradigm of Goudbeek and Krahmer (in press) with a dual task procedure developed by Kellogg, Olive, and Piolat (2007). In this dual task, participants perform a primary task while they simultaneously have to remember one of two simple stimuli that were presented. After finishing (part of) the main task, the same stimulus or a different one is presented and the participant has to indicate whether the sequence is one of two similar or two different stimuli. In their study, Kellogg et al. (2007) found differential effects of linguistic stimuli and visual stimuli on verbal and visual working memory. Based on this finding, we chose simple and similar words as stimuli for this task.

Figure 2 depicts the alternation between the working memory task and the interactive alignment task. The right side of Figure 2 shows the memory task, and the left side shows the interactive alignment paradigm. In the working memory task participant are visually presented with

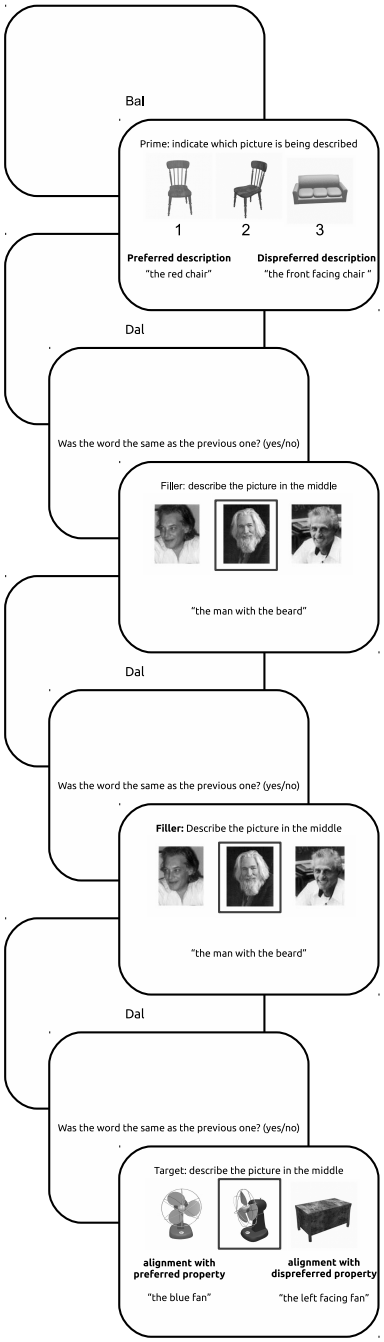

Figure 2: The sequence of events in a trial: the dual task (right frames) interlaced with the interactive alignment paradigm (leftmost frame).

one of two monosyllabic words, in our case "Bal" (English: ball) or "Dal" (English: valley) and have to keep this word in memory for a future

[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.

same/different comparison, after which the word needs to be updated in memory (Kellogg and colleagues used the syllables Ba and Da). Between the presentation of the word and the same/different judgment, participants identify and describe objects as in the original paradigm. Just like in the study by Goudbeek and Krahmer (in press), we presented the participants with descriptions using preferred or dispreferred attributes (to study content selection) or both (to study overspecification).

In each experiment, there were 20 critical trials and 20 filler items. Both the critical trials and fillers consisted of a prime, two fillers and a target (see the right side of Figure 2. In the selection experiment, 10 of the critical primes contained preferred attributes and 10 of the critical primes contained dispreferred attributes. In the overspecification experiment, all 20 critical primes contained overspecified referential expressions. Both the primes and the targets of the fillers were referred to by their type (e.g., *the chair*) or by pictures taken from the people domain that could be described as, for example, *the man with the beard*.

Both experiments had a complete within-subject design; all participants were exposed to all conditions. For the content selection experiment, this was operationalized as two blocks (one with preferred descriptions of furniture items and one with dispreferred descriptions of furniture), that were counterbalanced across participants. In the overspecification experiment, all 20 primes were presented to all participants.

After the experiment, we assessed whether our dual-task manipulation was successful by asking the participants whether they considered the task to be difficult or easy.

Results

Based on their answers on the difficulty questionnaire, we calculated the percentage of participants that found the task hard and those who found the task easy. Table 1 shows that participants in both dual task conditions found the task much more difficult compared to the participants in the single task condition.

Figure 3 summarizes the results for the content selection and the overspecification dual task experiment and compares these with the single task experiments from Goudbeek and Krahmer (in press). In the selection experiment, alignment occurs when participants use dispreferred attributes when primed with these, in the overspecification experiment, alignment occurs when participants produce overspecified descriptions following overspecified primes. The figure clearly shows that participants are less likely to align under load, and consequently more often only use the preferred attribute.

	Single task		**Dual task**	
	Easy	Hard	Easy	Hard
Selection	20/20	0/20	3/26	23/26
Overspec.	28/28	0/28	4/28	24/28
Total	100%	0%	17%	83%

Table 1: The reported difficulty for the single and dual task experiment for both experiments.

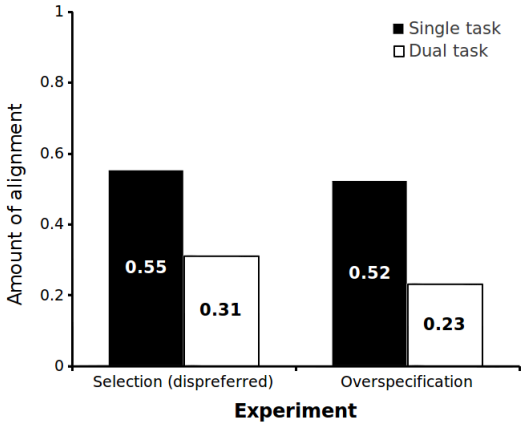

Figure 3: The amount of alignment for dispreferred primes and overspecified primes in the single task and dual task experiment.

We ran two analyses of variance (one for each experiment) with Task (single, dual) as independent variable, and amount of alignment as the dependent variable. These analyses confirm that participants aligned significantly less with dispreferred and with overspecified primes when they had to perform a secondary task ($F_{dispreferred}$ [1,46] = 5.76, $p < 0.02$, $\eta^2 = 0.12$; $F_{overspecification}$ [1,56] = 9.18, $p < 0.004$, $\eta^2 = 0.15$). Next, we tested whether the use of dispreferred attributes or overspecified references was more than zero, the value predicted by most REG algorithms, with the notable exception of the

model by Gatt et al. (2011). We found that the dispreferred attribute was used significantly more than zero ($t_{dispreferred}$ [25] = 5.49, $p < 0.01$) and participants produced significantly more overspecified references than predicted by 'traditional' REG algorithms ($t_{overspecification}$ [27] = 3.70, $p < 0.01$) when they had been previously presented with overspecified descriptions.

Discussion

This study investigated the effect of a demanding secondary task on the relative contribution of preference-based and alignment-based processes in the production of referring expressions. The references of participants were primed with preferred, dispreferred, and overspecified attributes. When the participants had to refer to a critical target picture, they could always use either a preferred, a dispreferred attribute, or both. The results of both experiments described in this paper show that when speakers have to refer under load, they are more likely to rely on their stable property preferences and less likely to align with their conversation partners than in the single task experiment.

Our interpretation of the experiment and the results favors an effect of task load on working memory capacity. The subsequent decrease in working memory capacity in turn hampers access of utterances from the dialogue context into working memory, as depicted in the right-hand process of the model proposed by Gatt and colleagues (Figure 1). However, this interpretation needs to be accompanied by a caveat, since the current experimental set-up and results do not exclude the possibility that the increased task load deteriorates processing at the perceptual level. After all, the continuous attention of our participants to a visual display, while simultaneously keeping either "Bal" or "Dal" in working memory could also negatively influence the ability to correctly perceive or remember the information presented in the prime. While valid, this interpretation becomes less evident when taking into account that the participants in both the single and dual task experiment hardly made any errors in indicating which picture was being described and in correctly referring to the target picture. This suggests that the dual task, while clearly harder than the single task, did not influence the performance of participants on the primary task.

Taken at face value, these results are in line with the work that assumes that adaptation is an effortful process that speakers must consciously activate (Horton & Keysar, 1996; Keysar et al., 1998). When working memory capacity is occupied by another task, there is insufficient capacity left for speaker to align with their dialogue partners. The results present a challenge for models that consider alignment processes to be automatic and relatively cost-free such as the interactive alignment model of Pickering and Garrod (Pickering & Garrod, 2004; Garrod & Pickering, 2004). After all, if aligning with a dialogue partner requires little mental effort, our participants would not be expected to stop aligning when faced with a dual task.

The results also support the model of Gatt et al. (2011) where dynamic alignment processes and stable property preferences are separate processes that work in parallel in reference production. In addition, they provide a first step to put relative weights on the two separate processes where stable and strong preferences will be favored over dynamic information that unfolds in dialogue when the participants are put under load.

Acknowledgments

The research reported in this paper forms part of the VICI project "Bridging the gap between psycholinguistics and Computational linguistics: the case of referring expressions", funded by the Netherlands Organization for Scientific Research (NWO grant 277-70-007). We thank Manon Yassa for assistance in running the experiments.

References

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.

Buschmeier, H., & Bergmann, K. (2009). An alignment-capable microplanner for natural language generation. In *In: Proceedings of the 12th european workshop on natural language generation (enlg* (pp. 82–89).

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the gen-

eration of referring expressions. *Cognitive Science*, *19*(2), 233-263.

Garrod, S., & Pickering, M. J. (2004, January). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Attribute preference and priming in reference production: Experimental evidence and computational modeling. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 2627 – 2632). Boston, USA.

Goudbeek, M., & Krahmer, E. (2010). Preferences versus adaptation during referring expression generation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 55–59). Uppsala, Sweden.

Goudbeek, M., & Krahmer, E. (in press). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Sciences*, *XX*, XX–XX.

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

Kellogg, R., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*(1), 382 –397.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1 –20.

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.*

Krahmer, E., & van Deemter, K. (in press). Computational generation of referring expressions: A survey. *Computational Linguistics*, *XX*, XX–XX.

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

Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, *27*, 169-226.

Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th annual conference of the cognitive science society (cogsci)* (pp. 685–690).

van Deemter, K., Gatt, A., van der Sluis, I., & Power, R. (in press). Generation of referring expressions: assessing the incremental algorithm. *Cognitive Science*, *XX*, XX–XX.