Attribute preference and priming in reference production
Experimental evidence and computational modeling

Albert Gatt (albert.gatt@um.edu.mt)
Institute of Linguistics, University of Malta
Tilburg Center for Cognition and Communication (TiCC), Tilburg University

Martijn Goudbeek (m.b.goudbeek@uvt.nl)
Tilburg Center for Cognition and Communication (TiCC), Tilburg University

Emiel Krahmer (e.j.krahmer@uvt.nl)
Tilburg Center for Cognition and Communication (TiCC), Tilburg University

Abstract

Referring expressions (such as *the red chair facing right*) often show evidence of *preferences* (Pechmann, 1989; Belke & Meyer, 2002), with some attributes (e.g. colour) being more frequent and more often included when they are not required, leading to *overspecified* references. This observation underlies many computational models of Referring Expression Generation, especially those influenced by Dale & Reiter's (1995) Incremental Algorithm. However, more recent work has shown that in interactive settings, priming can alter preferences. This paper provides further experimental evidence for these phenomena, and proposes a new computational model that incorporates both attribute preferences and priming effects. We show that the model provides an excellent match to human experimental data.

Keywords: Reference, production, Natural Language Generation, Computational Modeling

Introduction

In domains such as Figure 1, where a target referent needs to be distinguished from its distractors in context, people often produce *overspecified* descriptions such as *the red sofa facing right*, when a description containing fewer attributes would suffice (Pechmann, 1989; Eikmeyer & Ahlsèn, 1996; Belke & Meyer, 2002; Engelhardt, Bailey, & Ferreira, 2006). This finding challenges the assumption that speakers observe the Gricean Maxim of Quantity by not including any more information than is relevant for identification (cf. Olson, 1970, for an early adoption of this view).

One important observation in this regard is that certain attributes (for example, an object's colour), are more likely to be redundantly included in an overspecified description than others (such as size or orientation) (Pechmann, 1989; Belke & Meyer, 2002). The *preferred status* of such attributes may arise due to their perceptual salience, higher codability relative to other attributes (Belke & Meyer, 2002) and/or because they form an integral part of the conceptual representation of the object (Pechmann, 1989). On one interpretation of these findings, preferred attributes are selected first when a description is being formulated; since this is an incremental process, should later attributes be included which make them redundant, the whole description would be overspecified (Pechmann, 1989; Levelt, 1989).

This has important implications for computational models of referring expressions generation (REG), which seek

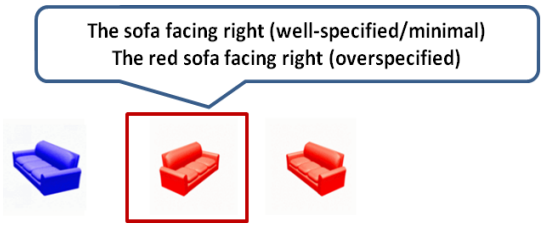


Figure 1: A referential domain

to model the process of *attribute selection* for identifying descriptions. Such models form an integral part of Natural Language Generation systems, which generate text or speech from non-linguistic input. Current REG models perform attribute selection primarily on the basis of discriminatory value: does a target attribute help to exclude some distractors in the domain? Some models (e.g. Dale, 1989; Gardent, 2002) seek to satisfy a strict interpretation of the Gricean maxim of quantity by selecting the smallest set of attributes that would uniquely identify the target referent(s). An alternative, more influential model is Dale and Reiter's (1995) Incremental Algorithm, which is in part inspired by the psycholinguistic literature and models attribute selection as an incremental search that prioritises more preferred attributes. As we show below, such models can overspecify in some situations. Furthermore, they have been shown to match speaker behaviour better than earlier models (Gatt, van der Sluis, & van Deemter, 2007; Gatt & Belz, 2010).

Many of the psycholinguistic studies cited above were undertaken in non-interactive settings, whereas recent psycholinguistic work on dialogue has highlighted the extent to which speakers's production choices are influenced by their interlocutors'. One aspect of this process, discussed by Clark *et al.* (Clark & Wilkes-Gibbs, 1986; Brennan. & Clark, 1996), is the 'negotiation' on the best way to refer to an object that characterises some interactive reference tasks. More recently, Pickering and Garrod (2004) have proposed the *Interactive Alignment* model, whereby interlocutors 'align' at various levels (for example, syntactic and semantic) as a result of a basic priming mechanism. There is substantial evidence that such priming occurs, particularly in interlocutors'

syntactic choices (e.g. Cleland & Pickering, 2003; Branigan, Pickering, McLean, & Cleland, 2007, among others). However, to our knowledge, there is less evidence for priming at a conceptual level.

One question that is particularly relevant from the point of view of the present paper is the extent to which alignment can influence attribute selection in reference, and how it interacts with preferences and overspecification. If priming occurs at a conceptual level, does hearing a description by another interlocutor modulate a speaker's attribute preferences? Recent work has suggested that speakers can indeed be primed to use non-preferred attributes (Goudbeek & Krahmer, 2010). One observation that arose from this work is that priming caused speakers to overspecify to a greater extent than would be predicted by a preference-based model. This raises the question of whether speakers can be *directly* primed to overspecify. From a computational point of view, an affirmative answer to this question would lend further support to the view that models such as the Incremental Algorithm, as well as related models that select attributes based purely on the basis of their discriminatory value (e.g. Dale, 1989), do not capture the full range of influences on referential choices that occur in interactive settings.

The present paper explores these questions further from both the experimental and the computational angles. After an overview of the Incremental Algorithm for REG, we describe the experiment by Goudbeek and Krahmer (2010) in more detail, and report on a new experiment using the same paradigm, which shows increased evidence for overspecification when overspecified primes are used. We then describe and evaluate a new computational model that seeks to incorporate both the classic findings on attribute preferences, and the novel findings on priming of dispreferred attributes and overspecification. We evaluate our model by comparing its output directly to the human descriptions elicited during these experiments.

Computational REG

Dale and Reiter's (1995) Incremental Algorithm (IA) has emerged as one of the most influential computational REG models. In searching for a distinguishing combination of attributes for a target referent, the IA uses a *preference order* to model preferences. For example, the attributes in Figure 1 could be ordered by preference as TYPE > COLOUR > ORIENTATION. To identify an intended referent r, the algorithm traverses the preference order, checking at each stage whether r's value on a given attribute excludes some distractors. The algorithm terminates when a referent has been fully distinguished, or when it runs out of attributes to choose from. For the target referent in the figure, the IA would not choose type (since all objects are sofas), but would choose colour (which excludes the blue sofa) and then orientation (which excludes the remaining sofa). This yields an overspecified description; ordering orientation before colour would have resulted in a minimal description, since orientation would have excluded both distractors immediately. Dale and Reiter also proposed a function to add type in case it is omitted by the search. Thus, this description could be realised as *the red sofa facing right*.

Overspecification in the IA occurs when, as a result of the preference order, an attribute (colour, in our example) is selected which excludes a set of distractors that is a proper subset of the distractors excluded by an attribute selected later (orientation). This behaviour is entirely deterministic, insofar as a preference order is pre-specified and cannot be overridden. On the other hand, the experimental work described earlier suggests that preferences can indeed be overridden through priming, which can also result in increased likelihood of overspecification. From a computational perspective, then, the question is how to incorporate preferences (for which robust evidence exists), while also introducing a sensitivity to context that can modulate them. We view preferences as a relatively stable phenomenon, related to an attribute's being inherently salient for a speaker. Modulation of such preferences as a result of priming might therefore occur as a result of a competing process, one which prioritises attributes that have been used earlier in an interaction, because it is cheaper to re-use conceptual material than to search for it anew. As we also show below, however, the priming/overspecification effects do not occur across the board; rather, they are best described as a (statistically significant) tendency. Thus, if a computational model is intended to match speaker behaviour, some degree of non-determinism will need to be introduced in the balancing act between preference-based and priming-based attribute selection.

Two experiments

The experiment by Goudbeek and Krahmer (2010, hereafter referred to as Experiment 1), which investigated the role of priming in the choice of preferred or dispreferred attributes, used the *Interactive Reference Understanding and Production* paradigm illustrated in Figure 2.

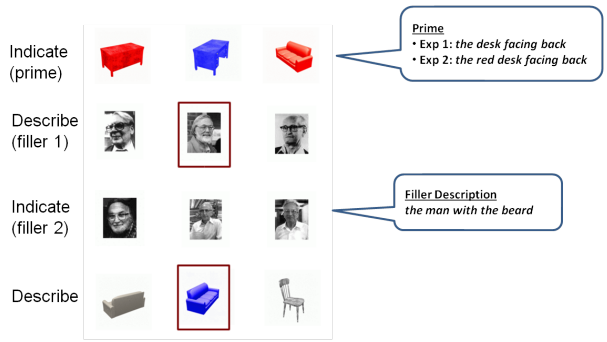
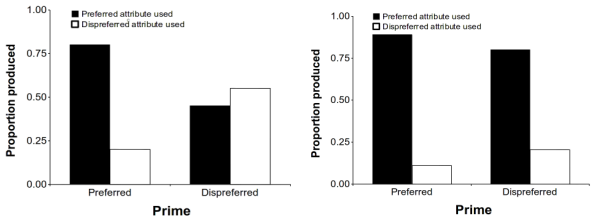


Figure 2: The Interactive Reference Understanding and Production paradigm

Participants were first asked to identify an object in a visual domain based on a pre-recorded description (marked as Exp 1 in the top panel of Figure 2), which contained either a *preferred* or a *dispreferred* attribute. This description, which



(a) Furniture (b) People

Figure 3: Alignment in Experiment 1

Table 1: Overspecification in the experiments (%)

	Experiment 1		Experiment 2
	Pref. Prime	Dispref. Prime	
Furniture	11.9	11.9	51.8
People	15.8	12.7	57.0
Overall	13.8	12.3	54.4

functioned as the prime, was never overspecified. Following two filler trials, during which participants first described and then identified objects in a different type of domain (e.g. people in Figure 2), they were asked to describe a target in the same domain as the prime. Crucially, the target could always be described using *either* the preferred *or* the dispreferred attribute; moreover, it had the same attributes as the one described in the prime, but different values (e.g. the prime target would have *back* for orientation, while the new target would have *front*).

The experiment was conducted on 26 Dutch speakers, using materials from the TUNA corpus (Gatt et al., 2007), a corpus of descriptions of objects in two domains (people and furniture). A Dutch version of this corpus has also been created (Koolen, Gatt, Goudbeek, & Krahmer, 2009), on the basis of which it was possible to determine which attributes were preferred (colour in the furniture domain; wearing glasses in the people domain) and which dispreferred (orientation in the furniture domain; having a tie in the people domain) by counting their frequencies. Participants described objects in both domains, with 20 preferred and 20 dispreferred primes in each, for a total of 80 trials.

The experiment sought to address two questions. The first concerned alignment, that is, whether the use of a dispreferred attribute in the prime (e.g. orientation, as in *facing back*) would increase the likelihood of a participant using the same attribute (though not the same value) in the critical trial (e.g. *facing front*). Note that a preference-based model such as the IA would never select a dispreferred attribute in this case, but would always return a description containing the preferred one.

The second question concerned overspecification. In the experimental domains, the IA would never produce an overspecified description, because the target referent can always be identified using only a preferred attribute. Hence, we asked whether people in this kind of situation overspecify to a greater extent than an algorithm such as the IA would predict.

As shown in Figure 3, participants showed strong evidence of alignment, with an increased tendency to use dispreferred attributes when they had been used in a prime description three turns earlier. A 2 (Domain) $\times$ 2 (Prime) within-subjects repeated measures ANOVA showed main effects of Domain ($F_{1,25} = 10.88; p = .01; \eta^2 = .3$) and Prime ($F_{1,25} = 6.43; p = 0.02; \eta^2 = 0.2$), as well as a significant interaction ($F_{1,25} = 5.74; p = .02; \eta^2 = .19$). The interaction is due to a greater tendency to use dispreferred attributes in the furniture, compared to the people domain. The left panel Table 1 also shows that people often overspecified by using both preferred and dispreferred attributes. A t-test showed that the rate of overspecification in both domains was significantly different from the base rate of 0% predicted by the IA (furniture: $t_{25} = 2.65, p = .01$; people: $t_{25} = 2.59, p = .02$).

These findings raise the question of whether overspecification can itself be primed. The new experiment reported here sought to test this directly, using the same experimental paradigm, but exposing participants to overspecified primes that contained both preferred and dispreferred attributes (marked as Exp 2 in Figure 2). The experiment was conducted on 28 Dutch speaking students from Tilburg University, none of whom had participated in Experiment 1, using the same materials and procedure, with the exception that the referring expressions used as primes were always overspecified. The right panel of Table 1 displays the proportion of overspecified descriptions produced by participants. The rate of overspecification rises dramatically in comparison to the rate observed in Experiment 1, suggesting that participants can indeed be primed to use both preferred and dispreferred attributes, and hence to overspecify. For the analysis, we combined these data with those from Experiment 1, using a mixed effects ANOVA with amount of overspecification as the dependent variable, domain as within-subjects variable and experiment (single prime or overspecified prime) as between-subjects variable. There was a significant effect of experiment ($F_{(1,52)} = 32.50, p < 0.001, \eta^2 = 0.36$), but no effect of domain and no interaction.

Thus, overspecified primes in Experiment 2 gave rise to more overspecified descriptions. This also strengthens one of the conclusions of Experiment 1, namely, that priming can result in increased use of dispreferred attributes (since overspecification in our experimental domains involve their use).

In summary, the experiments strongly support the view that both default attribute preferences and overspecification can be modulated by priming, thus challenging preference-based models such as the Incremental Algorithm. In the next section, we describe a new computational model that balances between attribute preferences and alignment effects, while also introducing a degree of non-determinism in attribute selection. The latter is crucial, for despite the statistically significant tendencies observed in these experiments, it is also clear that they do not constitute hard strategies. This stochastic behaviour is also a feature that distinguishes our model from fully deterministic REG algorithms such as the IA.

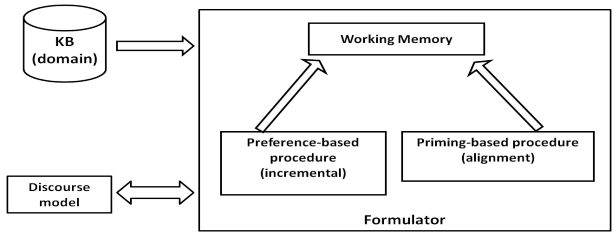


Figure 4: The parallel model

A computational model

We interpret the experimental findings as suggesting that there are two interacting forces – preferences and alignment – that influence attribute selection. Recall that in the experimental domains, a target referent could be distinguished using either a preferred or a dispreferred attribute. A preference-based model, such as the Incremental Algorithm (IA), would simply select the preferred attribute on the domains used in the experiments, but never the dispreferred one. Similarly, a priming-based procedure alone would select the most highly activated attribute, terminating immediately on finding that the attribute sufficed to distinguish the referent. If this were the case, we should observe 100% use of dispreferred attributes with dispreferred primes in Experiment 1, and around 50% across all trials in experiment 2. Both models would never overspecify on the experimental trials described above.

The model we propose, depicted in Figure 4, combines these processes in parallel, with both contributing to a working memory buffer. One way in which overspecification takes place is when the two processes contribute concurrently, resulting in two or more attributes in working memory that are both used in a description. The model's core is the Formulator module (the terminology is inspired by the model of Levelt, 1989, 1999) which is composed of the buffer and the two parallel processes. It makes use of a knowledge base (KB), which represents the domain (entities and their attributes and values), and a discourse model, which keeps a record of utterances spoken or heard so far.[1]

The preference-based procedure in the model is essentially a re-implementation of Dale and Reiter's (1995) Incremental Algorithm, with an attribute preference order determined using corpus frequencies, as in the experiments. The priming-based procedure is a spreading activation model, which works as follows. When a description is introduced into the discourse, the discourse model is updated, as a result of which the level of activation of attributes changes. Activation is estimated using an exponential decay function proposed by Buschmeier, Bergmann, and Kopp (2009), which combines *temporary* activation *ta* of an attribute A, which increases abruptly when an attribute is used and gradually decays to 0; and *permanent* activation *pa*, which increases when an attribute is used and maintains its level. These are shown in equations (1) and (2), where δ represents the time difference in milliseconds between the current and the previous usage of an attribute, and $f(A)$ is the frequency of A in the discourse. α and β are parameters determining the slopes of the functions; they are set to 2 in our simulations. The two functions are linearly combined to give an attribute's level of activation $act(A)$, as shown in equation (3), where v is a weight reflecting the relative importance of $pa(A)$ and $ta(A)$. This is set to 0.5 (i.e. equal weighting) in our experiments.

$$ta(A) = \exp\left(-\frac{\delta(A)-1}{\alpha}\right) \quad (1)$$

$$pa(A) = 1-\exp\left(-\frac{f(A)-1}{\beta}\right) \quad (2)$$

$$act(A) = v\cdot ta+(1-v)\cdot pa \quad (3)$$

A change in the discourse model causes all attribute activations to be updated. The upshot is that an attribute which has been used recently will increase abruptly in activation. Note that $ta(A)$ decreases with increasing $\delta(A)$, while $pa(A)$ increases with increasing $f(A)$. In line with our experimental findings, it is attributes that are activated, irrespective of their values. Thus, using an attribute like orientation in *the red desk facing back* will result in spreading activation to other values of orientation (e.g. *facing front*). The priming-based procedure then selects an attribute A for a new target referent if (a) $act(A)$ exceeds a threshold (empirically set to 0.4 in our simulations); (b) A has the highest activation of all the attributes of the referent.

As noted above, our experimental data suggests that a degree of non-determinism is at play in the interaction of the two processes. We model this using a *delay* parameter. The formulator schedules the two processes to run in parallel and calls each process at fixed intervals. Every call to a process results in its contributing an attribute to working memory, if (a) the buffer is not full to capacity; and (b) there are some attributes left to choose from. The interval at which each procedure is called is determined by the delay parameter: in our experiments, both procedures are assigned an equal delay (50ms), but this parameter functions as a ceiling. The actual delay is determined randomly at runtime as a value between 1 and the ceiling.

As the processes run in parallel, the formulator periodically checks the working memory buffer for new content, taking the attributes there and including them in the description, and emptying the buffer to free up working memory.[2] If the description is found to be distinguishing, the processes are terminated. For the purposes of our simulations, the working memory capacity was set to 2 (since this is the maximum

[1] A note on implementation: the model described here was implemented in Java, and exploits the multi-threading capacities of the Java Virtual Machine to schedule and run parallel processes.

[2] The checks made by the formulator are also randomly determined, based on the longest delay parameter of the two processes. In the simulations reported here, since both processes have an equal maximum delay of 50ms, this is randomly set at a value between 1 and 50.

number of attributes that can be selected for an object in our experimental domains).

This setup means that, on any given trial, one of the two processes may receive an advantage (because its delay is randomly determined to be lower than that of the other). As a result, it may contribute content to the working memory buffer before the other process, and potentially result in a non-overspecified description. On the other hand, there is also the possibility that both processes contribute before the buffer is checked and working memory is freed up once more. This is the principal way in which overspecification occurs. Another possibility is that a single process which happens to have a very brief delay contributes more than one attribute to working memory in succession, before the other process has time to contribute.

Table 2: Overspecification in the model simulations (%)

	Experiment 1		Experiment 2
	Pref. Prime	Dispref. Prime	
Furniture	13.8	20.0	49.1
People	13.8	16.5	45.7
Overall	13.8	18.3	44.4

Simulations

We evaluated the model against the experimental data from both experiments, focusing on the rate of overspecification. This is a particularly suitable metric for evaluation because it allows us to distinguish the model from the predictions of one based exclusively on preferences, or one based exclusively on priming; as observed above, both of these would predict a 0% rate in our experiments. However, it is important to emphasise that the model is intended as a more general characterisation of factors influencing attribute selection, with overspecification arising as a result of their interaction. Indeed, it is because of the different balance between preferences and priming in the two experiments that we expect different rates of overspecification in the two sets of trials. In Experiment 1, we expect our model to overspecify less. Since primes contain only a single attribute, only this one will exceed the activation threshold for the priming-based process. Of course, overspecification can still occur if the preference-based process selects another attribute concurrently. In the case of Experiment 2, both preferred and dispreferred attributes are primed equally, so that both have a chance of being selected by the priming-based procedure.

We performed a simulations for each of the two experiments, exposing the model to the same set of domains used with participants. In each case, the model referred to the same target referent, and the domain was set up to ensure that both preferred and dispreferred attributes could be used (i.e. were distinguishing). The model was primed by introducing a description into the discourse that contained either a preferred or a dispreferred attribute (for simulations of Experiment 1) or both (for simulations of Experiment 2). Since the model was run over the same trials as each participant, we can directly compare its rate of overspecification to that of humans, averaging over participants.

Table 2 displays proportions of overspecified descriptions produced by the model on the trials for each simulation. Note that overspecification occurs far less frequently for trials in Experiment 1 than Experiment 2. In Experiment 1, it also occurs more with dispreferred than preferred primes. This is primarily due to the priming-based process selecting the activated dispreferred attribute, while concurrently, the preference-based process selects a preferred attribute. By contrast, both attributes are primed in Experiment 2, making them equally likely to be selected at some point by the priming-based procedure. If a dispreferred attribute is selected concurrently with the selection of a preferred attribute by the preference-based process, the resulting description is overspecified.

For the simulations of both experiments, the model was statistically indistinguishable from humans, irrespective of domain (Experiment 1 simulation: ($t_{furniture}[25] = 1.07, t_{people}[25] = .16$. Experiment 2 simulation: $t_{furniture}[27] = .37, t_{people}[27] = 1.7$; all p's $> .1$).

For both experiments, these results diverge considerably from what a model based exclusively on preferences, such as the IA, or one based exclusively on priming, would predict.

Discussion and conclusions

This paper presented experimental evidence for the existence of multiple influences on attribute selection in reference, incorporating the findings in a model which matches human output very closely. The model takes an existing algorithm, the IA, as a starting point and views alignment as an interacting and competing force, modeled as a parallel, 'fast and frugal' strategy which is cheaper than the IA's preference-based search. The model thus distinguishes between *dynamic* effects arising in the course of an interaction, and more *stable* effects such as attribute preferences, which are likely to be related to properties of the human perceptual and conceptual apparatus (Pechmann, 1989).

An alternative model is conceivable, involving a single, dynamic search process whereby exposure to an attribute in an utterance directly alters the IA's preference order, by promoting the attribute to a higher position and making it more likely to be used later. However, this model would not only fail to make the distinction between relatively stable and relatively temporary factors, it would also fail to account for the different rates of overspecification observed in the experiments. Since only one attribute was required to distinguish a referent in the trials, this procedure would simply halt after selecting the first attribute and never overspecify. While the IA models overspecification exclusively as a result of redundant, but more preferred, attributes being selected before less preferred ones, our model acknowledges a second possible cause, as an effect of the 'interference' from a priming-based mechanism.

There are two questions raised by our work which are being addressed by current research. First, the model incorporates a

sizeable number of parameters. The delay parameter, which determines how parallel processes are scheduled at runtime, needs to be determined empirically, necessitating a detailed investigation of the time course of attribute selection based on preference and/or priming. Additionally, the use of a limited-capacity working memory buffer would predict that occupying the buffer would directly affect attribute selection. We are currently considering using dual-task paradigms, where participants carry out a memory task while performing a reference task. This may alter the rate of overspecification, in part because the dual task may cause participants to fall back on a 'cheap' strategy, relying exclusively on priming.

A second question concerns the status of the priming phenomenon itself. We have argued that our experiments show evidence of attribute-based (that is, conceptual or semantic) priming, in part because in Experiment 1, participants not only tended to re-use attributes they were primed with, but also overspecified by including information that was not in the prime. This suggests that participants were not merely re-using a syntactic template from the prime description. Nevertheless, the possibility remains that the priming mechanism is partially surface/syntactic, or strategic (in the sense that speakers were adopting a strategy for referring to furniture or people based on what they had heard). We are currently attempting to address these issues directly, by replicating these experiments using a bilingual priming paradigm, whereby the linguistic realisation of primes in one language is completely different from that of the descriptions uttered in a different language.

In summary, while the model proposed here matches human output, it also opens up a variety of new avenues for future research into referential strategies.

Acknowledgments

This work forms part of the project *Bridging the gap between psycholinguistics and computational linguistics: The case of Referring Expressions*, supported by the Netherlands Organization for Scientific Research (NWO).

References

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*(2), 237–266.

Branigan, H., Pickering, M., McLean, J., & Cleland, A. (2007). Participant role and syntactic alignment in dialogue. *Cognition*, *104*, 163-197.

Brennan., S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology*, *22*(6), 1482-1493.

Buschmeier, H., Bergmann, K., & Kopp, S. (2009). An alignment-capable microplanner for Natural Language Generation. In *Proc. 12th european workshop on natural language generation (ENLG'09)* (pp. 82–89).

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.

Cleland, A., & Pickering, M. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*, 214–230.

Dale, R. (1989). Cooking up referring expressions. In *Proc. 27th annual meeting of the association for computational linguistics (ACL'89).*

Dale, R., & Reiter, E. (1995). Computational interpretation of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(8), 233–263.

Eikmeyer, H. J., & Ahlsèn, E. (1996). The cognitive process of referring to an object: A comparative study of German and Swedish. In *Proc. 16th scandinavian conference on linguistics.*

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

Gardent, C. (2002). Generating minimal definite descriptions. In *Proc. 40th annual meeting of the association for computational linguistics (ACL'02).*

Gatt, A., & Belz, A. (2010). Introducing shared task evaluation to nlg: The TUNA shared task evaluation challenges. In E. Krahmer & M. Theune (Eds.), *Empirical methods in natural language generation.* Berlin and Heidelberg: Springer.

Gatt, A., van der Sluis, I., & van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proc. 11th european workshop on natural language generation (ENLG'07).*

Goudbeek, M., & Krahmer, E. (2010). Preferences versus adaptation during referring expression generation. In *Proc. 48th annual meeting of the association for computational linguistics (ACL'10).* (pp. 55–59).

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2009). Need I say more? On factors causing referential overspecification. In *Proc. workshop on production of referring expressions: Bridging computational and psycholinguistic approaches (*PRE-COGSCI'09*).*

Levelt, W. (1989). *Speaking: From intention to articulation.* MIT Press.

Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.

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

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.