The Use of Spatial Relations in Referring Expression Generation

Jette Viethen
Centre for Language Technology
Macquarie University
Sydney, Australia
jviethen@ics.mq.edu.au

Robert Dale
Centre for Language Technology
Macquarie University
Sydney, Australia
rdale@ics.mq.edu.au

Abstract

There is a prevailing assumption in the literature on referring expression generation that relations are used in descriptions only 'as a last resort', typically on the basis that including the second entity in the relation introduces an additional cognitive load for either speaker or hearer. In this paper, we describe an experiemt that attempts to test this assumption; we determine that, even in simple scenes where the use of relations is not strictly required in order to identify an entity, relations are in fact often used. We draw some conclusions as to what this means for the development of algorithms for the generation of referring expressions.

1 Introduction

In recent years, researchers working on referring expression generation have increasingly moved towards collecting their own data on the human production of referring expressions (REs) (Krahmer and Theune, 2002; van der Sluis and Krahmer, 2004; Gatt and van Deemter, 2006; Belz and Varges, 2007); and the recent Attribute Selection in the Generation of Referring Expressions (ASGRE) Challenge used the TUNA corpus (Gatt et al., 2007), which is the most extensive collection of referring expressions to date. While there is a substantial body of experimental work in psycholinguistics that looks at the human production of referring expressions (see, amongst more recent work, (Clark and Wilkes-Gibbs, 1986; Stevenson, 2002; Haywood et al., 2003; Jordan and Walker, 2005)) the large range of factors that play a role in language production mean that it is often the case that the specific question that one is interested in has not been studied before. So, NLG researchers have tended towards data gathering exercises that explore some specific aspect of referring expression generation, focussing on hypotheses relevant to algorithm development.

This paper is in the same mold. We are particularly interested in how people use spatial relations in referring expressions, and so in this paper we describe an experiment that explores the generation of relational referring expressions in a simple scene. Section 2 elaborates on our reasons for exploring this aspect of reference. Section 3 describes the experiment and provides some discussion of the results: our primary conclusion is that the assumption in the literature that relations are used 'as a last resort' does not appear to hold; relations are often used, even in simple scenes, when they are not strictly required, and it is likely that they would be more heavily used in more complex real-world scenes. We conclude in Section 4 with some observations as to how the results presented here might impact on the development of algorithms for referring expression generation, and outline some future work.

2 Spatial Relations in Referring Expression Generation

The bulk of the existing literature on referring expression generation (see, for example, Dale (1989), Dale and Reiter (1995), van Deemter (2006), Horacek (2004), Gatt and van Deemter (2006)) generally focuses on the use of non-relational properties, which can either be absolute (for example, colour) or relative (for example, size). We are interested in the

use of relational expressions, and in particular the use of spatial relations; the contexts of use we are interested in are task-specific, where, for example, we might want an omniscient domestic agent to tell us where we have placed a lost object (*You left your keys under the folder on the desk* ...), or to identify a hearer-new object in a cluttered scene (*the magazine at the bottom of the pile of papers next to the lampshade in the corner*). To develop agents with these kinds of referential capabilities, we want to acquire data that will inform the development of algorithms, either by automatically checking their ability to replicate the corpus, or as a baseline for assessing the performance of humans in an identification task based on the output of these algorithms.

In this paper, we describe an experiment that looks at how and when people use spatial relations in a simple scene. More specifically, we aim to explore the hypothesis that relations are always dispreferred over non-relational properties. This hypothesis appears to underly most approaches to referring expression generation that handle relations:

Gardent (2002) adopts a constraint based approach to deal with relations specifically geared at generating referring expressions that are as short as possible. As including a relation in a referring expression always entails the additional mention of at least a head noun for the related object, this approach inherently prefers properties over relations.

Krahmer and Theune (2002) extend the Incremental Algorithm (IA; Dale and Reiter (1995)) to handle relations. This requires a preference list over all properties and relations to be specified in advance. They explicitly choose to put spatial relations right at the end of that preference list, on the basis that 'It seems an acceptable assumption that people prefer to describe an object in terms of simple properties, and only shift to relations when properties do not suffice [...] it takes less effort to consider and describe only one object'.

As the referents in Varges' 2005 domain are all points on a map distinguishable only by their spatial relations to other objects, he has no choice but to use relations. However, he also adopts brevity as a main criterion for choosing which spatial relations to use.

Kelleher and Kruijff (2005, 2006) cite Clark and Wilkes-Gibbs' (1986) Principle of Minimal Cooperative Effort and Dale and Reiter's (1995) Principle of Sensitivity, as well as van der Sluis and Krahmer's (2004) production study, to motivate the ordering over the types of properties that can be used by their system; accordingly, their system only includes spatial (and hence relational) information in a referring expression if it is not possible to construct a description from non-relational properties.

These approaches would appear to favour the production of referring expressions containing long sequences of non-relational properties when a single relational property might do the job. We are interested, then, in whether it really is the case that relational expressions are dispreferred, and in determining when they might in fact be preferred.

To date, we are not aware of any substantial data sets that would allow this question to be explored. Both the TUNA corpus (Gatt et al., 2007) and the Macquarie Drawer data (Viethen and Dale, 2006) contain too few relational descriptions to allow us to draw conclusions about any kind of patterns; the GREC corpus (Belz and Varges, 2007) is not concerned with content selection at all, but rather studies the form of referring expressions used over a whole text; i.e. the choice between fully descriptive NPs, reduced NPs, *one*-anaphora and pronouns.

There are a number of corpora resulting from experiments involving human participants which contain referring expressions, such as Brennan and Clark's (1996) collection of tangram descriptions, the HCRC Map Task Corpus (Thompson et al., 1993), the COCONUT corpus (Jordan and Walker, 2005), and Byron and Fosler-Lussier's (2006) OSU Quake corpus. However, these contain whole conversations between communicative partners cooperating on a task, making it difficult to factor out the impact of prior discourse context on the referring expressions used.

3 The Data Gathering Experiment

3.1 General overview

We conducted a web-based production experiment to elicit referring expressions describing singular objects in very simple scenes. The study was aimed at shedding light on the question of whether spatial relations are indeed as dispreferred as suggested by the literature in those situations where non-relational descriptions are possible.

The Desiderata section of the report from the Workshop on Shared Tasks and Comparative Evaluation in NLG (Paris et al., 2007) emphasises the difficulties inherent in evaluating NLG systems due to the context dependency of language production: the output appropriate for any given referring expression generation system entirely depends on the particular task being performed. The data gathered in this experiment is intended to inform the development and evaluation of algorithms for the production of one-shot, fully distinguishing descriptions of simple objects in 3-dimensional scenes. The experiment is focussed on the adequate use of spatial relations in referring expressions.

In designing the materials for the experiment, we were conscious of a number of factors which we might expect to have an influence on the use of spatial relations: the prominence of other properties such as colour and size (i.e. whether most objects are of the same or similar size and colour, so that none are very distinct from the point of view of their direct properties); how easy it is to distinguish the target from the other objects around it; how easy it is to identify the target in the scene without using any locational information; and the visual salience of other objects which could serve as relatees in relational descriptions.

3.2 Method

3.2.1 Participants

In total, 74 participants completed the experiment. They were recruited by emailing 120 native English speakers and asking them to pass on the call for participation to other native or fluent English speakers. This resulted in a range of participants from a wide variety of backgrounds and age groups; most participants were in their early or mid twenties.

One participant indicated they were colour-blind, and another requested that their data be discarded. The data for a further nine participants was excluded from the analysis for reasons outlined in Section 3.2.4 below. Of the remaining 63 participants, 34 were male and 29 were female.

3.2.2 Materials

The stimuli for this study consisted of 20 jpeg images of simple scenes generated using Google SketchUp. Each scene contained three objects; each

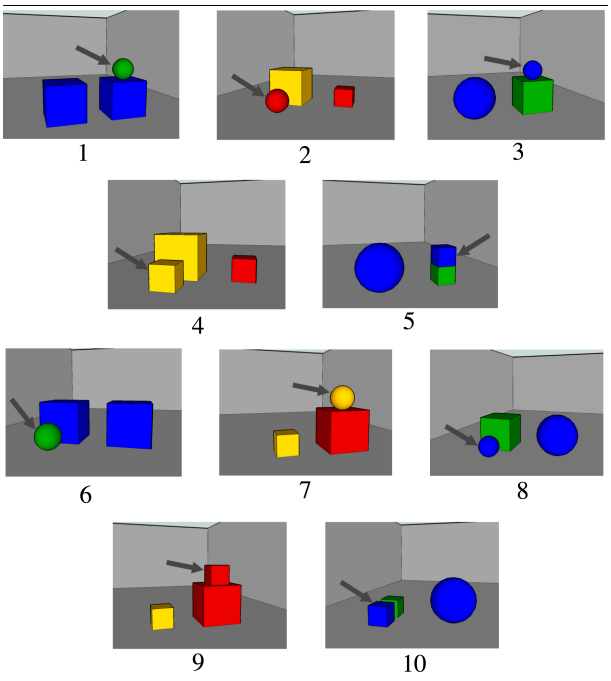

Figure 1: Trial Set 1: The five base configurations 1–5 and their counterparts 6–10 using the other type of target–landmark relation and orientation.

object was either a sphere or a cube. The objects could also be either large or small and were one of two colours; scenes either contained blue and green objects, or red and yellow objects. The *target object*, to be described by the participant, was marked by a grey arrow pointing to it; the target was always located either directly in front of or on top of one of the other two objects. We will refer to this other related object as the *landmark*, although there is of course no guarantee that participants actually included it into the description as the ground object in a spatial relation. The third object, which we refer to as the *distractor*, was located either to the left or the right of the target and landmark objects.

The 20 scenes are generated from five *base configurations*, differing in the type and size of the objects pictured. Figure 1 shows the five base configurations. They can be categorised by the length of the shortest possible description for the target object:

- in two of the base configurations it is possible to identify the target object using its type only;
- in one base configuration size alone would suf-

fice, although in line with past observations in the literature we would expect that type is always included as well;

- in one base configuration, colour and type are both necessary; and
- in the final base configuration, both size and colour are necessary, and again we would expect type to be included.

Importantly, there is no configuration in which the spatial relations between the objects are *required* in order to identify the target.

For each base configuration, we generated two scenes: in one scene, the target is located on top of the landmark object, and in the other, the target lies in front of the landmark. This allows us to investigate whether people prefer to use one type of spatial relation more than the other.

Five of the resulting 10 scenes were in the blue–green colour scheme, while the other five used red and yellow. The different colour schemes were an attempt to decrease the monotony of the task, so that we could show each participant more scenes. These 10 scenes, numbered 1 through 10, constituted our first trial set. A second trial set, with scenes numbered 11 through 20, was generated by producing the mirror image of each scene and using the opposite colour scheme. Mirroring the scenes had the same purpose as using the two different colour schemes. However, to be able to control any unwanted effect of these two variables we always used both variants.[1]

3.2.3 Procedure

On the experiment website, each participant was shown the scenes from one of the two trial sets in the order of the scene numbers. Under each scene, they had to complete the sentence *Please, pick up the ...* as if they were describing the object marked by the arrow to an onlooker.

To encourage the use of fully distinguishing referring expressions, participants were told that they had only one chance at describing the object. They were shown a sample scene for which they could provide an unrecorded (and unchecked) description. After being presented with all ten scenes in the trial, participants were asked to complete an exit questionnaire, which also gave them the option of having their data discarded, and asked for their opinion on whether the task became easier over time, and any other comments they might wish to make.

3.2.4 Data Processing

740 descriptions were elicited in the experiment. 10 of these were discarded in line with the participant's request, and 10 because the participant reported that they were colour-blind. After the data was cleaned and parsed, another 90 descriptions from 9 participants were discarded:

- One participant had consistently produced extremely long and complex descriptions using the ternary relation *between* and direct reference to the onlooker, the ground and parts of the objects: a typical example is *the red cube which rests on the ground and is between you and the yellow cube of equal size*. While these descriptions are interesting, in relation to the rest of the data they are such outliers that no real conclusions can be drawn from them.
- A further eight participants consistently used highly under-specified descriptions. We decided to discard the data from these participants since it seemed that they had not understood the need to provide a distinguishing description, rather than, for example, just indicating the type of the object.[2]

This resulted in a total of 630 referring expressions, with 30 for each scene in Trial Set 1 and 33 for each scene in Trial Set 2. We then applied some normalisation steps: the data was stripped of punctuation marks and other extraneous material (such as repetition of the *Please, pick up the*); in four cases, the dynamic spatial preposition *from* was deleted from descriptions such as *the green ball from on top of the blue cube*;[3] and spelling was normalised. The

[1] For brevity, where relevant we will use the form 'Scenes n+m' to refer to paired scenes across the two trial sets.

[2] Of course, underspecified descriptions are justified in many circumstances, and in real-life situations may even be necessary. However, the simple scenes used in this study do not fall into these classes.

[3] We are only interested in the static locative in these expressions; the use of the dynamic preposition is most likely due to the movement implied by the indicated picking-up action.

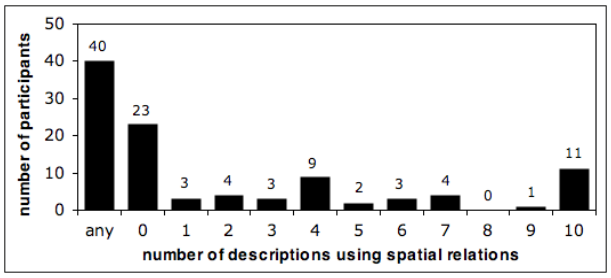

Figure 2: Number of participants who delivered n (0...10) relational descriptions.

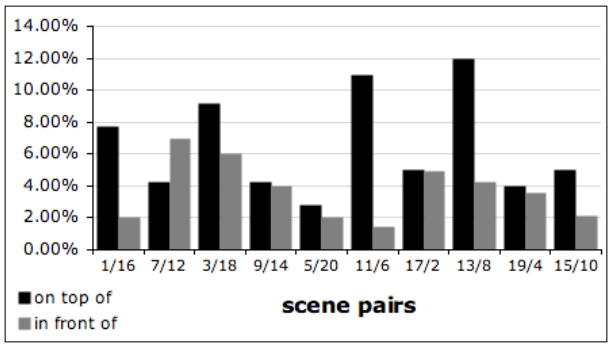

Figure 3: % of relational descriptions for each scene out of all relational descriptions produced by participants not using an exclusive strategy. Scenes are paired with their counterpart using the other target–landmark relation.

second object was stripped from comparatives such as *the smaller of the two green cubes* and converted to the form *the smaller green cube*, which in the context of our simple scenes is semantically equivalent.

3.3 Results

Over a third (231 or 36.6%) of the 630 descriptions in the resulting corpus use spatial relations despite the fact that relations were never necessary for the identification of the target. These 231 relational descriptions were produced by 40 (63.5%) of the 63 participants, while 23 (36.5%) of the participants never used spatial relations. This suggests that the use of relations is very much dependent on personal preference, a hypothesis that is further supported by the fact that 11 (i.e. over 25%) of the relation-using participants did so in *all* 10 referring expressions they delivered. Figure 2 shows the number of participants who produced exactly n descriptions containing at least one spatial relation, for n in the range $\{0 \ldots 10\}$.

From the above, we might hypothesise that some participants adopt a strategy of always using relational properties, and that others adopt a strategy of avoiding relational properties as much as possible. We further analysed the descriptions produced by participants who did *not* follow either of these two exclusive strategies to see how their choices varied across the different scenes; the spread is shown in Figure 3. Looking only at the descriptions produced by participants who sometimes, but not always, used spatial relations allows us to get a clearer view on which objects received most and least relational descriptions. This in turn affords an analysis of the impact the different features in the respective scenes have on the use of spatial relations.

41.7% of the remaining descriptions used relations. Interestingly, 63.6% of these relational descriptions were used for scenes where the target was located on top of the landmark object, while only 36.4% were from scenes where the target was in front of the landmark, suggesting that the use of the in-front-of relation may be relatively dispreferred.

Because the first scene always had the target on top of the landmark, this preference for using relational descriptions in on-top-of scenes might be due to a training effect that discourages people from using relations over time. However, if we do not take into account descriptions for the first 4 scenes of each trial set, this ratio is still large: 58.8% of the the remaining relational descriptions stemmed from scenes where the target was on top of the landmark, 41.2% of them from scenes with an on-top-of relation.

As expected, the orientation of the scenes and the colour scheme used did not have a significant impact on the use of spatial relations. For both these variables, the difference between values in use of relations was under 6 percentage points.

3.4 Discussion

We noted earlier that existing relation-handling referring expression generation algorithms generally disprefer relations and only add them to a description if absolutely necessary. This in essence mimics the behaviour of our participants who adopted the

exclusive Never-Use-Relations strategy.[4] These algorithms therefore only represent slightly more than one third of the participants in our study.

The analysis of the descriptions given by people who did not follow one of the two exclusive strategies indicates that the distribution of relational descriptions over the scenes is not random. In addition to modelling exclusive strategies, then, we may also want to capture in an algorithm the reasons why referring expressions for some scenes are more likely to include spatial relations than others.

In the remainder of this section we consider the conclusions that can be drawn from our data regarding the factors that impact on the choice of whether to use spatial relations in a referring expression.

Spatial Relations Are Used Even When Unnecessary: The main observation that can be made is that even in very simple scenes, where locatives are not necessary to distinguish the target from the other objects present, people show a tendency to use spatial relations to describe a target object to an onlooker. This contradicts the prevailing approach to the use of relations in referring expression generation. It is important to bear in mind that the scenes used in this study were extremely simple and could easily be taken in at one glance; it seems likely that when faced with a more crowded scene containing more complex objects, the tendency to incorporate possibly unnecessary spatial relations into descriptions would increase.

Training Effect: Note in Figure 4 that the targets in Scenes 1+11 received a disproportionally high number of descriptions containing spatial relations. While this fact could be attributed to the similar features of the two scenes (they only differed in orientation and colour scheme), it is much more likely that this is due to Scenes 1 and 11 being the first scenes of the respective trial sets. The drop-off in relational descriptions from beginning to end of the trial sets almost certainly results from a training effect, where people realised over time that relations were not necessary in any of the scenes. If we only consider the first two scenes in each trial set, where no training effect has taken hold, we find that 36 of the 58 (62.1%) descriptions for these scenes use spatial relations. The presence of some kind of training effect was also reported in the exit questionnaire by half of the participants.

[4] On the assumption that these participants would also resort to relations if they had to.

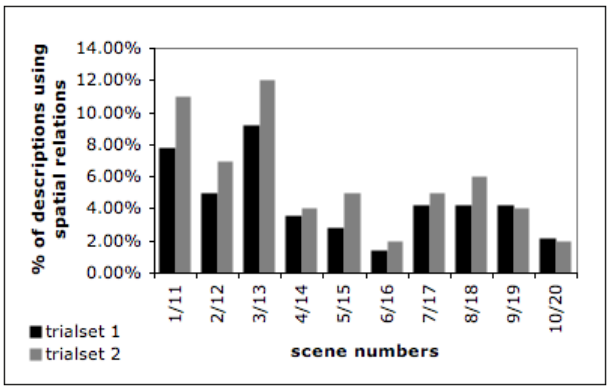

Figure 4: % of relational descriptions for each scene out of all relational descriptions produced by participants not using an exclusive strategy. Scenes are paired with their counterpart from the other trial set.

This training effect in itself is an interesting phenomenon. It suggests that people are much more likely to use spatial relations when they come anew to the task of identifying an object rather than when they are describing an object in a similar domain on a subsequent occasion.

Landmark Salience Encourages Use of Relations: Figure 4 shows that the highest spike in usage of spatial relations was recorded for Scenes 3+13; interestingly another, although much less pronounced, peak occurs for their counterpart scenes only differing in the type of target–landmark relation, 8+18.

These peaks cannot be explained by the training effect; in fact, they seem to be running contrary to it, indicating that some other reasonably strong factors are prompting the use of relations in these scenes.

Scenes 3, 8, 13, and 18 are the only scenes in which the landmark object is distinguishable from both other objects only by its type (cube) or its colour (see Figure 1). In addition, in each case the landmark is large resulting in high visual salience for the landmark. This in turn makes the relation to the landmark a salient feature of the target. The salience of the relation then causes people to add it to an already distinguishing description or even to prefer it

over the use of absolute properties.

***on top of* Is Preferred over *in front of*:** Although these four scenes all share the same base set of objects, the usage of spatial relations is considerably higher for Scenes 3+13 than for 8+18. This could either be entirely due to the training effect, but may also be influenced by the only difference between these two scenes: in Scenes 3+13, the target sits on top of the landmark, while in Scenes 8+18 it is lying in front of the landmark. The overall comparison of the data for scenes featuring an on-top-of relation with that for scenes with an in-front-of relation suggests that this also is a factor. Even if we only take into account Scenes 5–10 and 15–20, where we might expect the effect of training to have stabilised, people were almost one and a half times more likely to use a relation in a scene where the target was on top of rather than in front of the landmark (30 vs. 21 of the 111 relational descriptions for those scenes from people not using an exclusive strategy).

This finding is in accordance with Kelleher and Kruijff's (2006) approach of preferring topological spatial relations over projective ones. The semantics of projective spatial relations, such as *in front of*, depend on a frame of reference defining directions from some origin (in this case the landmark object), while topological relations, such as *on top of*, are semantically defined by relations such as intersection, containedness, and contiguity, and pose a lighter cognitive load on both discourse partners (see Tenbrink (2005) for an overview).

The impact of landmark salience and the preference for the on-top-of relation can also explain the low use of spatial relations for Scenes 4+14, 6+16 and 10+20 (see Figure 1). In these scenes it is very hard or even impossible to distinguish the landmark from the other objects using only non-relational properties, and the target is located in front of rather than on top of it. The possibility of describing the target in Scenes 6+16 only by its type or colour may be the reason for the extremely low usage of spatial relations in these scenes.

4 Conclusions

4.1 Consequences for Algorithm Development

We noted above that some participants adopted a Never-Use-Relations strategy, and some adopted an Always-Use-Relations stratgy. This might be modelled by the use of a parameter akin to the the Risky/Cautious distinction proposed by Carletta (1992) in her work on references in the Map Task corpus. The effect of this parameter in the context of the Incremental Algorithm would be to put spatial relations either at the front or at the end of the preference list of properties; this would ensure that they are either considered first for inclusion into a referring expression, or only when the other properties of the target do not suffice.

A more interesting problem is how to model the apparent preference of our participants to use relations in some scenes more than in others. Following our discussion above, the factors that lead to this preference seem to include the folllowing:

- the ease with which a potential landmark can be distinguished from the other objects in the scene;
- the visual salience of a potential landmark (in our case its size);
- the type of spatial relation between the target and a potential landmark; and
- the ease with which the target can be described without the use of spatial relations.

The visual salience of the target object is likely to also play a role; however, this was not tested in the current study, since all target objects were small.

Factors like these can be incorporated into a referring expression generation algorithm by taking them into account in the step that calculates which property of the target object should next be considered for inclusion in the referring expression. Instead of using a static preference list over all possible domain properties, a preference score for each property needs to be determined 'at run time'. Such a dynamic approach would also allow the consideration of the discourse salience of a property (perhaps due to its recent use in another referring expression), as well as the consideration that some properties are more likely to be used in combination with other specific properties. An example of this phenomenon is the combination of the property *hair-colour* with either *has-hair* or *has-beard* in the TUNA data. If *hair-colour* is included in a referring expression, at

least one of the other two properties is present as well.

The preference scores of the properties in a referring expression under construction would then combine into an adequacy score for the overall description, similar to Edmonds' (1994) concept of the speaker's confidence that a referring expression suffices for the communicative task at hand.

4.2 Future Work

As a next step, we aim to run experiments to separately confirm the impact that each of the different factors listed in Section 3.4 has on the use of spatial relations in referring expressions. In parallel, we will evaluate the human-produced descriptions in task-based evaluation schemes to assess whether the use of relations in certain categories of scenes is advantageous for an onlooker trying to identify the object that is being referred to.

Ultimately, the aim of this research is to develop an algorithm that incorporates the findings from both types of studies into the generation of referring expressions. Such an algorithm should not simply mimic the behaviour that our participants have displayed during the production experiment, but also take into account the findings of the task-based study, to ensure both naturalness and usefulness for the listener.

References

Anja Belz and Sebastian Varges. 2007. Generation of repeated references to discourse entities. In *Proceedings of the 11th European Workshop on Natural Language Generation*, pages 9–16.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1482–1493.

Donna K. Byron and Eric Fosler-Lussier. 2006. The OSU Quake 2004 corpus of two-party situated problem-solving dialogs. In *Proceedings of the 15th Language Resources and Evaluation Conference*.

Jean C. Carletta. 1992. *Risk-taking and Recovery in Task-Oriented Dialogue*. Ph.D. thesis, University of Edinburgh.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

Robert Dale. 1989. Cooking up referring expressions. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, BC.

Philip G. Edmonds. 1994. Collaboration on reference to objects that are not mutually known. In *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto, Japan.

Claire Gardent. 2002. Generating minimal definite descriptions. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, USA.

Albert Gatt and Kees van Deemter. 2006. Conceptual coherence in the generation of referring expressions. In *Proceedings of the 21st COLING and the 44th ACL Conference*, Sydney, Australia.

Albert Gatt, Ielka van der Sluis, and Kees van Deemter. 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the 11th European Workshop on Natural Language Generation*, pages 49–56.

Sarah Haywood, Martin J. Pickering, and Holly P. Branigan. 2003. Co-operation and co-ordination in the production of noun phrases. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, pages 533–538, Boston, MA.

Helmut Horacek. 2004. On referring to sets of objects naturally. In *Proceedings of the 3rd International Conference on Natural Language Generation*, pages 70–79, Brockenhurst, UK.

Pamela W. Jordan and Marilyn A. Walker. 2005. Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research*, 24:157–194.

John Kelleher and Geert-Jan M. Kruijff. 2005. A context-dependent model of proximity in physically situated environments. In *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Colchester, U.K.

John Kelleher and Geert-Jan M. Kruijff. 2006. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st COLING and the 44th ACL Conference*, Sydney, Australia.

Emiel Krahmer and Mariët Theune. 2002. Efficient context-sensitive generation of referring expressions. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, pages 223–264. CSLI Publications, Stanford, CA.

Cécile Paris, Donia Scott, Nancy Green, Kathy McCoy, and David McDonald. 2007. Desiderata for evaluation of natural language generation. In Robert Dale and Michael White, editors, *Proceedings of the Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*, pages 9–15, Arlington, VA.

Rosemary Stevenson. 2002. The role of salience in the production of referring expressions: A psycholinguistic perspective. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. CSLI, Stanford.

Thora Tenbrink. 2005. Semantics and application of spatial dimensional terms in English and German. Technical Report Series of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, No. 004-03/2005, Universities of Bremen and Freiburg, Germany.

Henry S. Thompson, Anne Anderson, Ellen Gurman Bard, Gwyneth Doherty-Sneddon, Alison Newlands, and Cathy Sotillo. 1993. The HCRC map task corpus: natural dialogue for speech recognition. In *Proceedings of the 1993 Workshop on Human Language Technology*, pages 25–30, Princeton, New Jersey.

Kees van Deemter. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics*, 32(2):195–222.

Ielka van der Sluis and Emiel Krahmer. 2004. The influence of target size and distance on the production of speech and gesture in multimodal referring expressions. In *Proceedings of the 8th International Conference on Spoken Language Processing (INTERSPEECH 2004)*, Jeju, Korea.

Sebastian Varges. 2005. Spatial descriptions as referring expressions in the maptask domain. In *Proceedings of the 10th European Workshop On Natural Language Generation*, Aberdeen, UK.

Jette Viethen and Robert Dale. 2006. Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of the 4th International Conference on Natural Language Generation*, pages 63–70, Sydney, Australia, July.