Generating Relational References: What Makes a Difference?

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

When we describe an object in order to enable a listener to identify it, we often do so by indicating the location of that object with respect to other objects in a scene. This requires the use of a *relational referring expression*; while these are very common, they are relatively unexplored in work on referring expression generation. In this paper, we describe an experiment in which we gathered data on how humans use relational referring expressions in simple scenes, with the aim of identifying the factors that make a difference to the ways in which humans construct referring expressions.

1 Introduction

The generation of referring expressions—the process of determining how best to describe an object in order to enable the hearer to identify it—is a widely explored topic in natural language generation research. We expect the first practical applications of the algorithms developed in this area to be those where the location of objects within scenes is a requirement: examples of such scenarios are the description of entities such as buildings and other landmarks in automatically-generated route descriptions (see Dale et al. (2005)) and the description of locations in 'omniscient room' scenarios, where an intelligent agent might try to tell you where you left your RFID-tagged keys. In these scenarios, it is very likely that the referring expressions generated will need to make use of the spatial relationships that hold between the intended referent and other entities in the domain; but, surprisingly, the generation of relational references is a relatively unexplored task. The few algorithms that address this task (Dale and Haddock (1991), Gardent (2002), Krahmer and Theune (2002), Varges (2005), Kelleher and Kruiff (2005, 2006)) typically adopt fairly simple approaches: they only consider spatial relations if it is not possible to fully distinguish the target referent from the surrounding objects in any other way, or they treat them in exactly the same as non-relational properties. As acknowledged by some of this work, this creates additional problems such as infinite regress and the inclusion of relations without regard for the properties of the landmarks that are associated with them.

To be able to develop algorithms that meet the requirements of applications like those just mentioned, we first need to have a better understanding of how humans use relational referring expressions. In this paper, we provide a detailed analysis of a set of experimental data that was gathered in a context where human subjects were encouraged to use relational expressions in order to identify intended referents.

In Section 2 we describe our data gathering experiment in some detail, explaining the rationale behind our development of the test data we used. In Section 3 we provide a summary of the corpus of referring expressions generated by the experiment. Section 4 provides a detailed analysis of the relational referring expressions in the corpus, identifying a range of phenomena of interest. Section 5 concludes the paper by drawing out some desiderata for the development of referring expression generation algorithms that follow from these observations.

2 The Data Gathering Experiment

2.1 General Overview

The real contexts in which referring expressions are used can be very complex. Consider the following hypothetical references in the motivating scenarios we used in the introduction:

(1)   Turn left after the second shopfront that has a 'For lease' sign in the window.

(2)   Your keys are under the loose leaf folder on the desk in the upstairs study.

In such real life situations, there are generally too many variables to permit carefully controlled experiments that would allow us to derive general principles for content determination. In line with almost all work in this area (see, for example, Brennan and Clark (1996), Thompson et al. (1993), Gorniak and Roy (2004), Jordan and Walker (2005), Byron and Fosler-Lussier (2006)), we therefore begin our explorations with very much simpler scenarios that allow us to explore specific hypotheses and to characterise the general strategies that humans seem to adopt; we can then apply these strategies in more complex scenes to see whether they continue to be applicable.

Our goal is to determine what characteristics of scenes impact on the use of spatial relations. The data gathering experiment we conducted had the form of a self-paced on-line language production study. Participants visited a website, where they first saw an introductory page with a set of simple instructions and a sample stimulus scene. Each participant was assigned one of two trial sets of ten scenes each. The scenes were presented successively in a preset order. Below each scene, the participant had to complete the sentence *Please pick up the . . .* in a text box before clicking on a button to see the next scene. The task was to describe the target referent in the scene (marked by a grey arrow) in a way that would enable a friend looking at the same scene to pick it out from the other objects.

74 participants completed the experiment. They were recruited by emailing self-reported native English speakers directly and asking them to pass on the invitation for participation. The participants were from a variety of different backgrounds and ages, but were mostly university-educated and in their early or mid twenties. For reasons outlined in Section 3.1, the data of 11 participants was discarded. Of the remaining 63 participants, 29 were female, while 34 were male.

2.2 Stimulus Design

2.2.1 The Components of Scenes

In order to explore even the most basic hypotheses with respect to the use of relational expressions, we require scenes which contain at least three objects. One of these is the intended referent, which we refer to as the *target*. The subject has to describe the target in such a way as to distinguish it from the other two objects in the scene. Although our scenes are such that spatial relations are never *necessary* to distinguish the target, the scenes are set up so that one of the two non-target objects was clearly closer to the target. We call this object the (potential) *landmark*; the third object in the scene is then the *distractor*.

To minimise the number of variables in our experiments, we restrict ourselves to only two kinds of objects, cubes and balls. The objects also vary in two dimensions: colour (either green, blue, yellow, or red); and size (either large or small).

To reduce the number of factors in our scene design, the landmark and distractor are always placed clearly side by side, and the target is located on top of or directly in front of the landmark. This results in four possible spatial configurations:

1. target on top of landmark, distractor to the left;

2. target on top of landmark, distractor to the right;

3. target in front of landmark, distractor to the left; and

4. target in front of landmark, distractor to the right.

Whether the distractor is to the left or to the right of the other two objects determines what we call the *orientation* of the scene.

2.2.2 Creating a Balanced Stimulus Set

Together, the apparently simple factors described above allow 16,384 distinct scenes to be constucted. Clearly this is too many for us to experimentally determine generalisable observations, so we took additional measures to reduce this number while keeping the stimulus set as balanced as possible.

Firstly, we decided to only use two colours in each scene, with the result that, in any scene, at least

two objects will have the same colour. The literature on visual salience (for an overview, see Chapters 1 and 2 of Pashler (1998), Yantis and Egeth (1999), and Caduff and Timpf (2007)) suggests that colour salience is relative rather than absolute. Consequently, we did not expect the individual colours to influence which objects were included in referring expressions; rather, we would expect that this is influenced by how different the colour of each object is from the colours of the other objects. We therefore attempted to keep the brightness of the different colours in each scene as constant as possible by choosing one of two colour 'templates' for each scene: blue+green and red+yellow. In order to make the experiment less monotonous for subjects, half of the scenes use one colour template, and the other half use the other.

Similarly, we would not expect the *orientation* of the scene to have an impact on the use of relations; so again we switch the orientation in half of the scenes in order to reduce monotony for the subject. However, having both orientations and both colour templates in the stimulus set still allows us to test for any unexpected impact of these factors.

Secondly, we chose to make all landmark objects cubes, simply because it might look unnatural to see an object balanced on top of a perfectly spherical ball. We also decided to exclude variation in the size of the target object from our analysis, making all target objects small; this avoids having to deal with the complexity of situations where the target object might obscure a smaller object which could otherwise be used as a landmark.

From the outset we had excluded situations in which spatial relations would be necessary to fully distinguish the target, i.e. scenes in which the target is identical to one or both of the other objects.

Even with these constraints, we still have 244 possible scenes. This is still too many to allow each possible scene to be tested a reasonable number of times without overloading our subjects; our aim, therefore, was to arrive at 20 stimulus scenes that could be divided into two equivalent trial sets, so that each participant would only have to describe 10 objects.

2.2.3 Scene Schemata

Our next step was to create five *schemata* (see Figure 1) as a basis for our final stimulus set. A schema
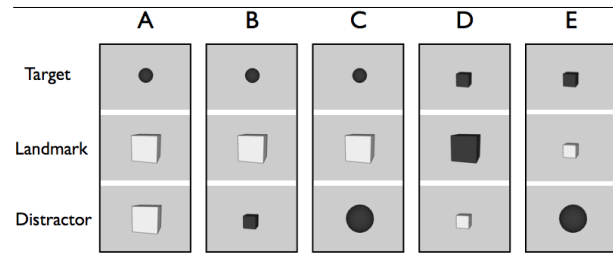

Figure 1: The *schemata* which form the basis for the stimulus scenes.

determines the type and size of each object in the scenes that are based on it, and determines which objects share colour. So, for example, in scenes based on Schema C, the target is a small ball; the landmark is a large cube; the landmark has a different colour from the target; and the distractor is a large ball sharing its colour with the target.

The design of these schemata is informed by three initial questions that we would like to explore:

1. Is the decision to use a spatial relation impacted by the length of the *minimal non-relational description* for the target?

A minimal non-relational description is the shortest referring expression that uniquely describes the target referent without using any relations to other objects or to the scene as a whole. To test this we need to have scenes where the target can be distinguished only by its type (Schemata A and B), scenes where a combination of type with either colour or size suffices to describe the target (Schemata C and E), and scenes where all three non-relational properties are necessary (Schema D). This question is examined in Section 4.3.1.

2. Is the use of spatial relations impacted by the similarity between target and landmark?

In Schemata A, B and C, these two objects share no properties; and in Schemata D and E, they share two properties, which is as similar as they can be without sharing all their properties. We look at the influence of target–landmark similarity in Section 4.3.2.

3. Is the use of spatial relations impacted by the similarity between landmark and distractor?

Based on the assumption that the salience of the landmark might have an impact, and that its salience is in turn determined by its similarity to the other
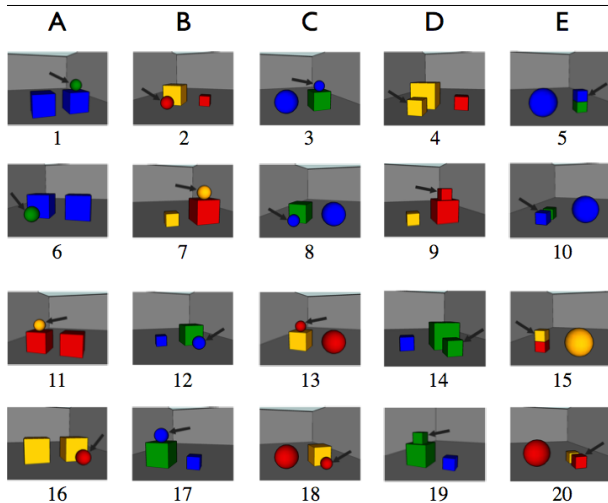
Figure 2: The stimulus scenes. The letters indicate which schema from Figure 1 each column of scenes is based on.

two objects, we were mindful that the stimuli should be also somewhat balanced regarding the landmark's similarity to the distractor. In Schema A, landmark and distractor are identical; in Schema E, they are completely distinct from each other; and in Schemata B, C and D, they share only one property value (their size in Schema C, and their type in Schemata B and D). The effect of landmark–distractor similarity is examined in Section 4.3.3.

2.2.4 Deriving Scenes from Schemata

From each schema we generated four scenes, resulting in the 20 stimulus scenes shown in Figure 2.

First we created Scenes 1–5, each based on a different one of the five schemata, by alternating between the colour templates and the spatial relations between target and landmark and the orientation of the scene, i.e. the position of the distractor.

We then created Scenes 6–10 by changing the spatial relation and orientation in each of the first five scenes. For example, Scene 8 was generated from Scene 3 by placing the target in front instead of on top of the landmark and flipping the scene, so that the distractor is on the right instead of the left. Scenes 1–10 constitute Trial Set 1.

The second trial set containing Scenes 11–20 was generated from the first one by changing the colour template and again flipping each scene along the vertical middle axis.

Because all 20 stimuli were generated from the same five schemata, they naturally fall into five different *groupings*. Due to the systematic generation process, we ensured that the target–landmark relation, the orientation and the colour template of the scenes within each *grouping* never fully coincide: if two scenes share one characteristic (e.g. the colour template), then they differ on the other two (in that case, orientation and target–landmark relation).

3 The GRE3D3 Corpus[1]

3.1 Data Normalisation and Filtering

One of the 74 participants asked for their data to be discarded. We also disregarded the data of one other participant who reported to be colour-blind. One participant consistently produced very long and syntactically complex referring expressions including reference to parts of objects and the onlooker, such as *the red cube which rests on the ground and is between you and the yellow cube of equal size*. While these descriptions are very interesting, they are clearly outliers in our data set.

Eight participants consistently only used type to describe the target object, for example simply typing *cube* for the target in Scene 5. These descriptions were excluded from the corpus under the assumption that the participants had not understood the instructions correctly or were not willing to spend the time required to type fully distinguishing referring expressions for each trial. After removal of this data, we have 630 descriptions: 30 for each of the ten scenes from Trial Set 1, and 33 for each scene in Trial Set 2.

Before conducting any quantitative data analysis, we carried out some syntactic and lexical normalisation. In particular, we corrected spelling mistakes; normalised names for *colour* values and head nouns (such as *box* instead of *cube*); and replaced complex syntactic structures such as relative clauses with semantically equivalent simpler ones such as adjectives. These normalisation steps should be of no consequence to our analysis, as we are solely inter-

---

[1]We refer to the data set resulting from the experiment described above as the GRE3D3 Corpus; the name stands for 'Generation of Referring Expressions in 3D scenes with 3 Objects'. The corpus is available online at http://www.ics.mq.edu.au/~jviethen/spatial/index.html.

ested in exploring the semantic content of referring expressions, not their lexical and syntactic surface structure.

3.2 Object Properties Used in the Corpus

The non-relational properties that occur in the corpus to describe the target object are limited to the expected type, colour and size. In five cases size was used as a relation between two objects in the form of *the same size as* or *larger than*. The spatial relations used are also mainly the expected on-top-of and in-front-of relations; however, left-of and right-of were used four and six times respectively, to relate the distractor object to one of the other two.

In addition to these direct and spatial relational properties, some participants used locative expressions such as *to the left*, *to the right*, *in the front* and in the top to describe the objects. Strictly speaking, these express relations to regions of the scene; however, we treat them as a distinct class of properties, which allows us to study the use of spatial relations between two objects separately from the use of relations to the scene.[2]

4 Data Analysis

In this section, we first provide a general analysis of the GRE3D3 Corpus data in Section 4.1. Then, in Section 4.2, we look at how characteristics of the overall scene impact on the forms of reference that people choose to use; and in Section 4.3, we look at how similarities between the objects in the scene impact on the forms of reference chosen.

4.1 General Analysis

Despite the fact that spatial information was not necessary in any of the trials for full object identification, 224 (35.6%) of the 630 descriptions contained a spatial relation to the landmark. 10 descriptions also contain a spatial relation to the distractor object. An example of a relational description that was given for the target in Scene 1 is *the green ball on top of the blue cube*.

Colour is used in 497 (78.9%) of all scenes. In 141 of these it is used twice, once for the target and once for the landmark; and in the ten descriptions

that make reference to the distractor, it is mentioned for all three objects. Size is used considerably less frequently, in only 288 (45.7%) of all descriptions. Only 43 descriptions use size for both target and landmark; and five mention size for all three objects.

62 relations to regions of the scene were found in the corpus, as in *the blue cube in the front* for Scene 10. However, 74.2% of these locative expressions described the landmark and not the target referent itself. Four descriptions contained two relations to regions of the scene. Note that we concentrate our analysis here on spatial relations between actual objects and exclude relations to regions of the scene.

As noted in (Viethen and Dale, 2008), some participants used relations for all 10 scenes presented to them, and some never used relations. Not counting relations to regions of the scene, 9 participants opted to always use relations, and 24 adopted a relation-free strategy.

Figure 3 shows a falling trend in the use of relations from the first scenes the participants saw to the later ones. On the basis of subjects' comments provided on completion of the experiment, we believe that this is a kind of 'laziness effect', whereby subjects noticed after a few trials that relations were unnecessary and stopped using them. This suggests that the use of relations would potentially be even higher than one third in a setting where no such laziness effect could occur, such as the first mention of an object in a real-world situation.

4.2 The Impact of Scene Characteristics on Referring Expressions

4.2.1 Target–Landmark Relation

Recall that the target can either be on top of the landmark or in front of it, with half of our scenes (the odd-numbered ones) being of the first type, and half (the even-numbered ones) being of the second type.

In (Viethen and Dale, 2008) we observed that the number of relational descriptions in the GRE3D3 Corpus is much higher for scenes with an on-top-of relation than for those with in-front-of relations between target and landmark: 63.6% of relational descriptions are used in on-top-of scenes, while only 36.4% are used in scenes where the target is in front of the landmark.

However, in that earlier analysis, we counted the

---

[2]Note that in (Viethen and Dale, 2008) relations to regions of the scene were treated the same as relations to other objects.
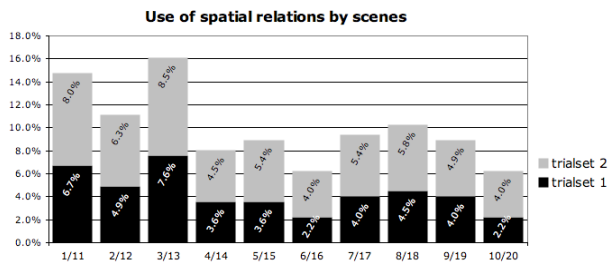
Figure 3: The use of relations for each scene. Scenes that only differ in colour template and scene orientation are stacked.

use of all locative expressions, including those involved in reference to parts of the scene as well as to landmar objects. Excluding such confounding data from the analysis confirms that if people chose to use a spatial relation, they did in fact use on-top-of relations in 'on-top-of' scenes, and in-front-of relations in 'in-front-of' scenes. This investigation also verifies the preference for on-top-of over in-front-of. Overall, 58.0% of the 224 relational descriptions contain an on-top-of relation between target and landmark, while only 42.0% use in-front-of relations, which is a statistically significant difference (χ^2=8.98, df=2, $p < .05$). Expressed differently, this means that of the 315 trials where the scene contained an on-top-of relation between target and landmark, on-top-of was used in 41.2% of the instances, and in-front-of was only used in 29.8% of the 315 'in-front-of' trials (χ^2=5.79, df=2, $p < .05$).

This uneven distribution of spatial information between the two types of target–landmark relations could be due to the fact that, in 'in-front-of' scenes, the landmark is partially occluded by the target object. Another factor that might be at play is people's general preference for relations on the frontal axis of a landmark over those on the lateral axis (see Tenbrink (2005), p. 18).

4.2.2 Colour Templates and Scene Orientation

As expected, the colour template used in a scene did not have a significant effect on whether relations between objects were used, whether colour was used, or whether a relation to a region of the scene was used; in each case, the difference between the two colour templates is one percentage point or less.

Similarly, whether the distractor was displayed to

the left or the right of the landmark had no significant effect on the content of the referring expressions produced.

4.3 The Impact of Object Similarity on Referring Expressions

As mentioned in Section 2.2, the psychological literature on visual salience suggests that the visual salience of an object is mainly determined by how much its appearance differs from that of its surrounding objects. We attempted to keep other factors, such as the distance between objects and occlusion of objects, as constant as possible between scenes, so that an object becomes more salient as the number of properties it shares with the other two objects decreases, and as the similarity of those other two objects to each other increases. We therefore report below results from the analysis of the extent to which the similarity between target, landmark and distractor influences the use of relations.

For the analysis of reference behaviour in the simple scenes we used, it is sufficient to adopt this rough notion of visual salience. However, it is not clear how it would carry across to more complex scenes with more properties and more objects. For example, it is not obvious whether an object that shares its type with all other objects, but is unique in all other properties, is more salient than an object that is unique in type but shares all other properties with one object each.

4.3.1 Target Salience

In order to answer Question 1 from Section 2.2.2 (whether the use of relations is impacted by the length of the *minimal non-relational description* for the target), we test for the influence of the number of properties the target shares with any other object in the scene as a rough measure of its visual salience.

In Schemata A and B, the target never shares its type with one of the other objects, which means it can be described uniquely by using only type. In Schemata C and E, it shares two properties with another object, so the length of its minimal description is two. While not being identical with either object in Schema D, here the target shares all three property values with one of the other objects and therefore can only be distinguished by a referring expression of at least length three.

The use of relations to the landmark drops slightly for targets that are very similar, i.e. less visually salient, from 36.9% to 30.2%. Although this drop is not statistically significant, the trend is surprising. If the trend is confirmed in future experiments, it might be explained by a hypothesis similar to that proposed by Edmonds (1994): the information giver's confidence has to be high enough to satisfy them that they have conveyed sufficient information. In addition to the requirement that the description has to be fully distinguishing, having to meet such a confidence threshold might compel people to add visually salient information, such as spatial relations to prominent landmarks, to a short referring expression, while the threshold might already be satisfied by longer non-relational descriptions.

4.3.2 Target–Landmark Similarity

Since we excluded scenes where the landmark is a ball, the test for influence of target type coincides with the one that can help us answer Question 2 from Section 2.2.2: Is the use of relations impacted by the similarity between target and landmark? The way in which we have defined visual salience means that the similarity between two objects impacts the visual salience of both in the same way. We expected that a visually salient target would receive fewer relational descriptions, while a visually salient landmark would result in relations being used more often, so this test can also give us an idea of which object's visual salience yields more influence on the use of relations. In scenes based on Schemata D and E, target and landmark are similar, while they are completely distinct in those based on Schemata A, B and C.

This factor had a clear effect on the use of all properties in the corpus: For scenes with dissimilar target and landmark, the use of spatial relations was at 1.4 times significantly higher (40.2% of all descriptions for these scenes) than for the scenes where they were similar (28.6%) (χ^2=8.94, df=1, $p < .01$); the use of the non-relational properties (colour and size), on the other hand, was much lower. This outcome, supported by the non-conclusive results of the previous section, suggests that, at least in our domain, the visual salience of the landmark has more impact on the choice of whether a relation between it and the target gets included in the referring expres-

sion than the visual salience of the target itself.

4.3.3 Landmark Salience

If the visual difference between the objects in a scene is the main factor determining whether and which spatial information gets used, then it stands to reason that the visual salience of a potential landmark (i.e., how different it is from the surrounding objects) has a particularly high influence on whether the specific relation between it and the target object gets used. In (Viethen and Dale, 2008) we made a weak claim to this effect based on the fact that particularly many relational descriptions were recorded for scenes based on Schema C, where the landmark differs from the other two objects in type and colour. More detailed analysis of the data reveals that, indeed, the more different the landmark is from the other two objects, the more likely it is to be included in the referring expression via a spatial relation.

In 46.8% of all trials where the scene contained a landmark very dissimilar to both other objects (Schema C), a spatial relation to the landmark was used. In scenes where the landmark was more similar to the distractor than the target (Schemata A and B), this number fell to 36.9%. Landmarks sharing more properties with target than distractor (Schemata D and E) were included in referring expressions in only 28.6% of trials where they occurred. The difference between these three categories is statistically significant at χ^2=12.55 (df=2, $p < .01$).

Considering that the target object is already in focus when the landmark's salience is evaluated, the visual difference between the landmark and the target might be of more importance. This is in line with the findings in the previous section.

Below we look at three individual factors that influence the landmark's visual salience; the difference of its size and colour from the other two objects and its overall difference from the distractor in all properties.

Landmark Size Interestingly, the percentage of relational descriptions is significantly lower for scenes where the landmark has a unique size (33.3%) than for those where it shares its size with the distractor object (42.1%). For scenes where the landmark has the same size as the target, the num-

ber is significantly lower again (27.0%), as expected (χ^2=9.24, df=2, $p < .01$). However, these numbers bear the influence of both the 'laziness effect' reducing relation use for scenes displayed late and the fact that in half of the trials where landmark and distractor had the same size, the landmark was unique in its two other properties, colour and type.

Landmark Colour   The trend for the influence of the landmark's difference in colour to the other objects is similar, although not as pronounced as for size and not statistically significant: landmarks with unique colour are used in spatial relations slightly less often (36.8%) than those sharing their colour with the distractor (37.3%), while spatial relations with landmarks sharing the colour of the target are only used in 30.2% of the trials where this occurred.

Interestingly, the use of the landmark's colour shows the same trend as the influence of colour difference on the use of a landmark: it is most likely to be actually mentioned in relational descriptions for scenes where it is the same as the colour of the target object (94.7%). In relational descriptions for scenes where the landmark had a unique colour, this colour was mentioned in only 73.4% of the cases. Landmark colour was only verbalised in 48.9% of the relational descriptions where it was the same as the colour of the distractor. These differences are statistically significant at χ^2=11.45 (df=2, $p < .01$). This data can help us determine which properties to mention for a landmark, once the decision to refer to landmark has been made.

Landmark–Distractor Similarity   We attempt to answer Question 3 from Section 2.2.2 by testing for an effect of the visual salience of the distractor on referring expressions. Visual salience here is again effectively the inverse of the number of properties shared with other objects in the scene.

The use of spatial relations is at its lowest for scenes where the distractor is completely distinct from the landmark object (Schema E). 27.0% of all descriptions for these scenes contained relations. While this difference is not statistically significant, we had expected an opposite trend due to the distinctness of the distractor from the landmark in these scenes making the target–landmark cluster appear as a more salient unit, compelling people to use the spatial relation between them. This was not the case;

but again this might be due to the influence of the laziness effect having set in.

There was almost no difference between the usage of relations for scenes where landmark and distractor are identical (based on Schema A) and those where they share one property value (based on Schemas B, C, D). When they were identical relations were used in 37.3%, otherwise in 37.8% of all scenes in either condition.

5 Conclusions

The data analysis presented above clearly suggests that all scenes are not the same when it comes to the use of spatial relations in referring expressions. It brings up the interesting problem of 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 be determined by the visual salience of the different objects involved. In particular, the visual salience of a potential landmark seems to be of high importance. The specific factors we found to have an impact include:

- the type of spatial relation that holds between the target and a potential landmark;

- the visual salience of the target as measured by the number of properties it has that are not shared with other objects around it;

- the visual salience of the landmark as measured by its inverse similarity to the target referent.

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. A preference score for each property needs to be determined 'at run time', which would also allow the consideration of the *discourse* salience of a property, which results from previous mentions and the purpose of the discourse. The preference scores of the properties in a referring expression under construction would then combine into an adequacy score for the overall description.

In future work we intend to embody these observations in an implemented algorithm for the generation of referring expressions.

References

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*.

David Caduff and Sabine Timpf. 2007. On the assessment of landmark salience for human navigation on the assessment of landmark salience for human navigation. *Cognitive Processes*.

Robert Dale and Nicolas Haddock. 1991. Content determination in the generation of referring expressions. *Computational Intelligence*, 7(4):252–265. Robert?

Robert Dale, Sabine Geldof, and Jean-Philip Prost. 2005. Using natural language generation in automatic route description. In *Journal of Research and Practice in Information Technology*, volume 37, pages 89–105.

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.

Peter Gorniak and Deb Roy. 2004. Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, 21:429–470.

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.

Harold Pashler. 1998. *Attention*. Psychology Press, Hove, UK.

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.

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. 2008. The use of spatial relations in referring expression generation. In *Proceedings of the 5th International Conference on Natural Language Generation*, Salt Fork OH, USA.

Steven Yantis and Howard E. Egeth. 1999. On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3):661–676.