

# Generation of Relative Referring Expressions based on Perceptual Grouping

**Kotaro FUNAKOSHI**

Department of Computer Science  
Tokyo Institute of Technology  
Meguro Ôokayama 2-12-1,  
Tokyo 152-8552, Japan  
koh@cl.cs.titech.ac.jp

**Naoko KURIYAMA**

Department of Human System Science  
Tokyo Institute of Technology  
Meguro Ôokayama 2-12-1,  
Tokyo 152-8552, Japan  
kuriyama@hum.titech.ac.jp

**Satoru WATANABE**

Department of Computer Science  
Tokyo Institute of Technology  
Meguro Ôokayama 2-12-1,  
Tokyo 152-8552, Japan  
satoru\_w@cl.cs.titech.ac.jp

**Takenobu TOKUNAGA**

Department of Computer Science  
Tokyo Institute of Technology  
Meguro Ôokayama 2-12-1,  
Tokyo 152-8552, Japan  
take@cl.cs.titech.ac.jp

## Abstract

Past work of generating referring expressions mainly utilized attributes of objects and binary relations between objects. However, such an approach does not work well when there is no distinctive attribute among objects. To overcome this limitation, this paper proposes a method utilizing the perceptual groups of objects and  $n$ -ary relations among them. The key is to identify groups of objects that are naturally recognized by humans. We conducted psychological experiments with 42 subjects to collect referring expressions in such situations, and built a generation algorithm based on the results. The evaluation using another 23 subjects showed that the proposed method could effectively generate proper referring expressions.

## 1 Introduction

In the last two decades, many researchers have studied the generation of referring expressions to enable computers to communicate with humans about concrete objects in the world.

For that purpose, most past work (Appelt, 1985; Dale and Haddock, 1991; Dale, 1992; Dale and Reiter, 1995; Heeman and Hirst, 1995; Horacek, 1997; Krahmer and Theune, 2002; van Deemter, 2002; Krahmer et al., 2003) makes use of attributes of an intended object (the target) and binary relations between the target and others (distractors) to distinguish the target from distractors. Therefore, these methods cannot generate proper referring expressions in situations where no significant surface difference exists between the target and distractors, and no binary relation is useful to distinguish the target. Here, a *proper* referring expression means a concise and natural linguistic expression enabling hearers to distinguish the target from distractors.

For example, consider indicating object  $b$  to person  $P$  in the situation shown in Figure 1. Note that person  $P$  does not share the label information such as  $a$  and  $b$  with the speaker. Because object  $b$  is not distinguishable from objects  $a$  or  $c$  by means of their appearance, one would try to use a binary relation between object  $b$  and the table, i.e., “A ball to the right of the table”.<sup>1</sup> However, “to the right of” is not a discriminatory relation, for objects  $a$  and  $c$  are also located to the right of the table. Using  $a$  and  $c$  as a reference object instead of the table does not make sense, since  $a$  and  $c$  cannot be uniquely identified because of the same reason that  $b$  cannot be identified. Such situations have never drawn much attention, but can occur easily and frequently in some domains such as object arrangement (Tanaka et al., 2004).

van der Sluis and Krahmer (2000) proposed using gestures such as pointing in situations like those shown in Figure 1. However, pointing and gazing are not always available depending on the positional relation between the speaker and the hearer.

In the situation shown in Figure 1, a speaker can indicate object  $b$  to person  $P$  with a simple expression “the front ball” without using any gesture. In order to generate such an expression, one must be able to recognize the salient perceptual group of the objects and use the  $n$ -ary relative relations in the group.<sup>2</sup>

In this paper, we propose a method of generat-

---

<sup>1</sup>In this paper, we simply assume that all participants share the appropriate reference frame (Levinson, 2003). We mention this issue in the last section.

<sup>2</sup>Although Krahmer *et al.* claim that their method can handle  $n$ -ary relations (Krahmer et al., 2003), they provide no details. We think their method cannot directly handle situations we discuss here.

ing referring expressions that utilizes  $n$ -ary relations among members of a group. Our method recognizes groups by using Thórisson’s algorithm (Thórisson, 1994). As the first step of our research project, we deal with the limited situations where only homogeneous objects are randomly arranged (see Figure 2). Therefore, we handle positional  $n$ -ary relation only, and other types of  $n$ -ary relation such as size, e.g., “the *biggest* one”, are not mentioned.

Speakers often refer to multiple groups in the course of referring to the target. In these cases, we can observe two types of relations: the *intra-group relation* such as “the front two *among* the five near the desk”, and the *inter-group relation* such as “the two *to the right of* the five”. We define that a subsumption relation between two groups is an intra-group relation.

In what follows, Section 2 explains the experiments conducted to collect expressions in which perceptual groups are used. The proposed method is described and evaluated in Section 3. In Section 4, we examine a possibility to predict the adequacy of an expression in terms of perceptual grouping. Finally, we conclude the paper in Section 5.

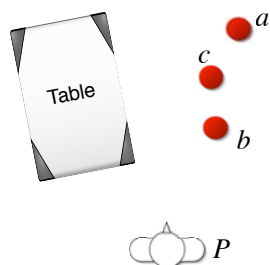


Figure 1: An example of problematic situations

## 2 Data Collection

We conducted a psychological experiment with 42 Japanese undergraduate students to collect referring expressions in which perceptual groups are used. In order to evaluate the collected expressions, we conducted another experiment with a different group of 44 Japanese undergraduate students. There is no overlap between the subjects of those two experiments. Details of this experiment are described in the following subsections.

### 2.1 Collecting Referring Expressions

**Method** Subjects were presented 2-dimensional bird’s-eye images in which several objects of the same color and the same size were arranged and the subjects were requested to convey a target object to the third person drawn in the same image. We used 12 images of arrangements. In each image, three

to nine objects were arranged manually so that the objects distributes non-uniformly. An example of images presented to subjects is shown in Figure 2. Labels  $a, \dots, f, x$  in the image are assigned for purposes of illustration and are not assigned in the actual images presented to the subjects. Each subject was asked to describe a command so that the person in the image picks a target object that is enclosed with dotted lines. When a subject could not think of a proper expression, she/he was allowed to abandon that arrangement and proceed to the next one. Referring expressions designating the target object were collected from these subjects’ commands.

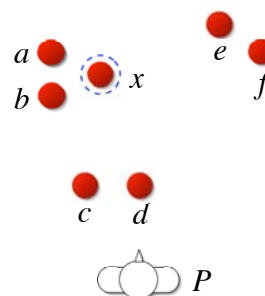


Figure 2: A visual stimulus of the experiment

**Analysis** We presented 12 arrangements to 42 subjects and obtained 476 referring expressions. Twenty eight judgments were abandoned in the experiment. Observing the collected expressions, we found that starting from a group with all of the objects, subjects generally narrow down the group to a singleton group that has the target object. Therefore, a referring expression can be formalized as a sequence of groups (SOG) reflecting the subject’s narrowing down process.

The following example shows an observed expression describing the target  $x$  in Figure 2 with the corresponding SOG representation below it.

“*hidari oku ni aru mittu no tama no uti no itiban migi no tama.*”  
(the rightmost ball among the three balls at the back left)

SOG:  $[\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}]$ <sup>3</sup>  
where  
 $\{a, b, c, d, e, f, x\}$  denotes all objects in the image (total set),  
 $\{a, b, x\}$  denotes the three objects at the back left, and  
 $\{x\}$  denotes the target.

<sup>3</sup>We denote an SOG representation by enclosing groups with square brackets.

Since narrowing down starts from the total set, the SOG representation starts with a set of all objects and ends with a singleton group with the target. Translating the collected referring expressions into the SOG representation enables us to abstract and classify the expressions. On average, we obtained about 40 expressions for each arrangement, and classified them into 8.4 different SOG representations.

Although there are two types of relations between groups as we mentioned in Section 1, the expressions using only intra-group relations made up about 70% of the total.

## 2.2 Evaluating the Collected Expressions

**Method** Subjects were presented expressions collected in the experiment described in Section 2.1 together with the corresponding images, and were requested to indicate objects referred to by the expressions. The presented images are the same as those used in the previous experiment except that there are no marks on the targets. At the same time, subjects were requested to express their confidence in selecting the target, and evaluate the conciseness, and the naturalness of the given expressions on a scale of 1 to 8.

Because the number of expressions that we could evaluate with subjects was limited, we chose a maximum of 10 frequent expressions for each arrangement. The expressions were chosen so that as many different SOG representations were included as possible. If an arrangement had SOGs less than 10, several expressions that had the same SOG but different surface realizations were chosen. The resultant 117 expressions were evaluated by 49 subjects. Each subject evaluated about 29.5 expressions.

**Analysis** Discarding incomplete answers, we obtained 1,429 evaluations in total. 12.2 evaluations were obtained for each expression on average.

We measured the quality of each expression in terms of an *evaluation value* that is defined in (1). This measure is used to analyze what kind of expressions are preferred and to set up a scoring function (6) for machine-generated expressions as described in Section 3.1.

$$\begin{aligned} & (\textit{evaluation value}) \\ = & (\textit{accuracy}) \times (\textit{confidence}) \\ \times & \frac{(\textit{naturalness}) + (\textit{conciseness})}{2} \quad (1) \end{aligned}$$

According to our analysis, the expressions with only intra-group relations (84 samples) obtained high accuracies (Ave. 79.3%) and high evaluation

values (Ave. 33.1), while the expressions with inter-group relations (33 samples) obtained lower accuracies (Ave. 69.1%) and lower evaluation values (Ave. 19.7).

The expressions with only intra-group relations are observed more than double as many as the expressions with inter-group relations. We provide a couple of example expressions indicating object  $x$  in Figure 2 to contrast those two types of expressions below.

- without inter-group relations
  - “the rightmost ball among the three balls at the back left”
- with inter-group relations
  - “the ball behind the two front balls”

In addition, expressions explicitly mentioning all the objects obtained lower evaluation values. Considering these observations, we built a generation algorithm using only intra-group relations and did not mention all the objects explicitly.

Among these expressions, we selected those with which the subjects successfully identified the target with more than 90% accuracy. These expressions are used to extract parameters of our generation algorithm in the following sections.

## 3 Generating Referring Expressions

### 3.1 Generation Algorithm

Given an arrangement of objects and a target, our algorithm generates referring expressions by the following three steps:

- Step 1:** enumerate perceptual groups based on the proximity between objects
- Step 2:** generate the SOG representations by combining the groups
- Step 3:** translate the SOG representations into linguistic expressions

In the rest of this section, we illustrate how these three steps generate referring expressions in the situation shown in Figure 2.

#### Step 1: Enumerating Perceptual Groups.

To generate perceptual groups from an arrangement, Thórisson’s algorithm (Thórisson, 1994) is adopted.

Given a list of objects in an arrangement, the algorithm generates groups based on the proximity of the objects and returns a list of groups. Only groups containing the target, that is  $x$ , are chosen because

$$\begin{aligned}
& \text{SOG: } [\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}] \\
& \rightarrow E(R(\{a, b, c, d, e, f, x\}, \{a, b, x\})) + E(\{a, b, x\}) + E(R(\{a, b, x\}, \{x\})) + E(\{x\}) \\
& \rightarrow \text{“hidari oku no”} + \text{“mittu no tama”} + \text{“no uti no migihasi no”} + \text{“tama”} \\
& \quad \text{(at the back left)} \quad \text{(three balls)} \quad \text{(rightmost . . . among)} \quad \text{(ball)}
\end{aligned}$$

Figure 3: An example of surface realization

we handle intra-group relations only as mentioned before, and that implies that all groups mentioned in an expression must include the target. Then, the groups are sorted in descending order of the group size. Finally a singleton group consisting of the target is added to the end of the list if such a group is missing in the list. The resultant group list,  $GL$ , is the output of Step 1.

For example, the algorithm recognizes the following groups given the arrangement shown in Figure 2:

$$\begin{aligned}
& \{\{a, b, c, d, e, f, x\}, \{a, b, c, d, x\}, \\
& \quad \{a, b, x\}, \{c, d\}, \{e, f\}\}.
\end{aligned}$$

After filtering out the groups without the target and adding a singleton group with the target, we obtain the following list:

$$\{\{a, b, c, d, e, f, x\}, \{a, b, c, d, x\}, \{a, b, x\}, \{x\}\}. \quad (2)$$

### Step 2: Generating the SOG Representations.

In this step, the SOG representations introduced in Section 2 are generated from the  $GL$  of Step 1, which generally has a form like (3), where  $G_i$  denotes a group, and  $G_0$  is a group of all the objects. Here, we narrow down the objects starting from the total set ( $G_0$ ) to the target ( $\{x\}$ ).

$$[G_0, G_1, \dots, G_{m-2}, \{x\}] \quad (3)$$

Given a group list  $GL$ , all possible SOGs are generated. From a group list of size  $m$ ,  $2^{m-2}$  SOG representations can be generated since  $G_0$  and  $\{x\}$  should be included in the SOG representation. For example, from a group list of  $\{G_0, G_1, G_2, \{x\}\}$ , we obtain four SOGs:  $[G_0, \{x\}]$ ,  $[G_0, G_1, \{x\}]$ ,  $[G_0, G_2, \{x\}]$ , and  $[G_0, G_1, G_2, \{x\}]$ .

For example, one of the SOG representations generated from list (2) is

$$[\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}]. \quad (4)$$

Note that any two groups  $G_i$  and  $G_j$  in a list of groups generated by Thórisson’s algorithm with regard to one feature, e.g., proximity in this paper, are

mutually disjoint ( $G_i \cap G_j = \phi$ ), otherwise one subsumes the other ( $G_i \subset G_j$  or  $G_j \subset G_i$ ). No intersecting groups without a subsumption relation are generated.

### Step 3: Generating Linguistic Expressions.

In the last step, the SOG representations are translated into linguistic expressions. Since Japanese is a head-final language, the order of linguistic expressions for groups are retained in the final linguistic expression for the SOG representation. That is, an SOG representation  $[G_0, G_1, \dots, G_{n-2}, \{x\}]$  can be realized as shown in (5), where  $E(X)$  denotes a linguistic expression for  $X$ ,  $R(X, Y)$  denotes a relation between  $X$  and  $Y$ , and ‘+’ is a string concatenation operator.

$$\begin{aligned}
& E(G_0) + E(R(G_0, G_1)) + E(G_1) + \dots \\
& \quad + E(R(G_{n-2}, \{x\})) + E(\{x\}) \quad (5)
\end{aligned}$$

As described in Section 2.2, expressions that explicitly mention all the objects obtain lower evaluation values, and expressions using intra-group relations obtain high evaluation values. Considering these observations, our algorithm does not use the linguistic expression corresponding to all the objects, that is  $E(G_0)$ , and only uses intra-group relations for  $R(X, Y)$ .

Possible expressions of  $X$  are collected from the experimental data in Section 2.1, and the first applicable expression is selected when realizing a linguistic expression for  $X$ , i.e.,  $E(X)$ . Therefore, this algorithm produces one linguistic expression for each SOG even though there are some other possible expressions.

For example, the SOG representation (4) is realized as shown in Figure 3.

Note that there is no mention of all the objects,  $\{a, b, c, d, e, f, x\}$ , in the linguistic expression.

### 3.2 Evaluation of Generated Expressions

We implemented the algorithm described in Section 3.1, and evaluated the output with 23 undergraduate students. The subjects were different from those of the previous experiments but were of the same age group, and the experimental environment

	Accuracy (%)	Naturalness	Conciseness	Confidence	Eval. val.
Human-12-all	87.3	4.82	5.27	6.14	29.3
Human-12-90	97.9	5.20	5.62	6.50	35.0
Human-12-100	100	5.36	5.73	6.65	37.2
System-12	91.0	5.60	6.25	6.32	40.1
System-20	88.4	5.09	5.65	6.25	35.2
System-Average	89.2	5.24	5.82	6.27	36.6

Table 1: Summary of evaluation

was the same. The evaluation of the output was performed in the same manner as that of Section 2.2.

The results are shown in Table 1. “Human-12-all” shows the average values of all expressions collected from humans with 12 arrangements as described in Section 2.2. “Human-12-90” and “Human-12-100” show the average values of expressions by humans that gained more than 90% and 100% in accuracy in the same evaluation experiment respectively.

“System-12” shows the average values of expressions generated by the algorithm for the 12 arrangements used in the data collection experiment described in Section 2.1. The algorithm generated 18 expressions for the 12 arrangements, which were presented to each subject in random order for evaluation.

“System-20” shows the average values of expressions generated by the algorithm for 20 randomly generated arrangements that generate at least two linguistic expressions each. The algorithm generated 48 expressions for these 20 arrangements, which were evaluated in the same manner as that of “System-12”.

“System-Average” shows the micro average of expressions of both “System-12” and “System-20”.

“Accuracy” shows the rates at which the subjects could identify the correct target objects from the given expressions. Comparing the accuracies of “Human-12-\*” and “System-12”, we find that the algorithm generates good expressions. Moreover, the algorithm is superior to human in terms of “Naturalness” and “Conciseness”. However, this result should be interpreted carefully. Further investigation of the expressions revealed that humans often sacrificed naturalness and conciseness in order to describe the target as precisely as possible for complex arrangements.

#### 4 Scoring SOG Representations

The algorithm presented in the previous section outputs several possible expressions. Therefore, we have to choose one of the expressions by calculating their scores.

The scores can be computed using various measures, such as complexity of expressions, and salience of referent objects. In this section, we investigate whether the adequacies of the courses of narrowing down can be predicted: that is, whether meaningful scores of SOG representations can be calculated.

##### 4.1 Method for SOG Scoring

An SOG representation has a form as stated in (3). We presumed that, when a speaker tries to narrow down an object group from  $G_i$  to  $G_{i+1}$ , there is an optimal ratio between the dimensions of  $G_i$  and  $G_{i+1}$ . In other words, narrowing down a group from a very big one to a very small one might cause hearers to become confused.

For example, consider the following two expressions that both indicate object  $x$  in Figure 2. Hearers would prefer (i) to (ii) though (ii) is simpler than (i).

- (i) “the rightmost ball among the three balls at the back left”
- (ii) “the fourth ball from the right”

In fact, we found (i) among the expressions collected in Section 2.1, but did not find (ii) among them. Our algorithm generated both (i) and (ii) in Section 3.2, and the two expressions gained the evaluation values of 44.4 and 32.1 respectively.

If our presumption is correct, we can expect to choose better expressions by choosing expressions that have adequate dimension ratios between groups.

##### Calculation Formula

The total score of an SOG representation is calculated by averaging the scores given by functions  $f_1$  and  $f_2$  whose parameters are dimension ratios between two consecutive groups as given in (6), where  $n$  is the number of groups in the SOG.

$$score(SOG) = \frac{1}{n-1} \left\{ \sum_{i=0}^{n-3} f_1 \left( \frac{dim(G_{i+1})}{dim(G_i)} \right) + f_2 \left( \frac{dim(\{x\})}{dim(G_{n-2})} \right) \right\} \quad (6)$$

The dimension of a group  $dim$  is defined as the average distance between the centroid of the group and that of each object. The dimension of the singleton group  $\{x\}$  is defined as a constant value. Because of this idiosyncrasy of the singleton group  $\{x\}$  compared to other groups,  $f_2$  was introduced separately from  $f_1$  even though both functions represent the same concept, as described below.

The optimal ratio between two groups, and that from a group to the target were found through the quadratic regression analysis of data collected in the experiment described in Section 2.2.  $f_1$  and  $f_2$  are the two regression curves found through analysis representing correlations between dimension ratios and values calculated based on human evaluation as in formula (1). We could not find direct correlations between dimension ratios and accuracies.

## 4.2 Results

We checked to what extent the scores of generated expressions given by formula (6) conformed with the human evaluation given by formula (1) as *agreement*. Agreement was calculated as follows using 20 randomly generated arrangements described in Section 3.2.

First, the generated expressions were ordered according to the score given by formula (6) and the human evaluation given by formula (1). All binary order relations between two expressions were extracted from these two ordered lists of expressions. The agreement was defined as the ratio of the same binary order relations among all binary order relations.

The agreement between scores and the human evaluation was 45.8%. The score did not predict SOG representations that would generate better expressions very well. Further research is required to conclusively rule out the use of dimension ratios for prediction or whether other factors are involved.

## 5 Concluding Remarks and Future Work

This paper proposed an algorithm that generates referring expressions using perceptual groups and  $n$ -ary relations among them. The algorithm was built on the basis of the analysis of expressions that were collected through psychological experiments. The performance of the algorithm was evaluated by 23 subjects and it generated promising results.

In the following, we look at future work to be done.

**Recognizing salient geometric formations:** Thórisson's algorithm (Thórisson, 1994) cannot recognize a linear arrangement of objects as a group, although such arrangements are quite salient

for humans. This is one of the reasons for the disconformity between the evaluations given by the algorithm and those of the human subjects.

We can enumerate most of such geometric arrangements salient for human subject by referring to geometric terms found in lexicons and thesauri such as "line", "circle", "square" and so on. Thórisson's algorithm should be extended to recognize these arrangements.

### Using relations other than positional relations:

In this paper, we focused on positional relations of perceptual groups. Other relations such as degree of color and size should be treated in the same manner.

Thórisson's original algorithm (Thórisson, 1994) takes into account these relations as well as positional relations of objects when calculating similarity between objects to generate groups. However, if we generate groups using multiple relations simultaneously, the assumption used in Step 1 of our algorithm that any pair of groups in an output list do not intersect without a subsumption relation cannot be held. Therefore, the mechanism generating SOG representations (Step 2 in Section 3.1) must be reconsidered.

### Resolving reference frames and differences of perspective:

We assumed that all participants in a conversation shared the same reference frame. However, when we apply our method to conversational agent systems, e.g., (Cavazza et al., 2002; Tanaka et al., 2004), reference frames must be properly determined each time to generate referring expressions. Although there are many studies concerning reference frames, e.g., (Clark, 1973; Herskovits, 1986; Levinson, 2003), little attention has been paid to how reference frames are determined in terms of the perceptual groups and their elements.

In addition to reference frames, differences of perspective also have to be taken into account to produce proper referring expressions since humans often view spatial relations between objects in a 3-dimensional space by projecting them on a 2-dimensional plane. In the experiments, we presented the subjects with 2-dimensional bird's-eye images. The result might have been different if we had used 3-dimensional images instead, because the projection changes the sizes of objects and spatial relations among them.

**Integration with conventional methods:** In this paper, we focused on a limited situation where inherent attributes of objects do not serve any identifying function, but this is not the case in general. An algorithm integrating conventional attribute-based methods and the proposed method should be formu-

lated to achieve the end goal.

A possible direction would be to enhance the algorithm proposed by Kraemer *et al.* (Kraemer *et al.*, 2003). They formalize an object arrangement (*scene*) as a labeled directed graph in which vertices model objects and edges model attributes and binary relations, and regard content selection as a subgraph construction problem. Their algorithm performs searches directed by a cost function on a graph to find a unique subgraph.

If we consider a perceptual group as an ordinary object as shown in Figure 4, their algorithm is applicable. It will be able to handle not only intra-group relations (e.g., the edges with labels “front”, “middle”, and “back” in Figure 4) but also inter-group relations (e.g., the edge from “Group 1” to “Table” in Figure 4). However, introducing perceptual groups as vertices makes it difficult to design the cost function. A well-designed cost function is indispensable for generating concise and comprehensible expressions. Otherwise, an expression like “a ball in front of a ball in front of a ball” for the situation shown in Figure 1 would be generated.

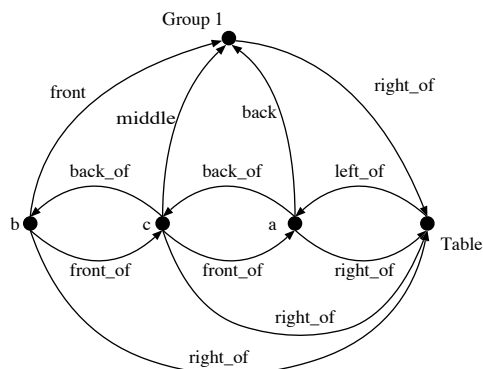


Figure 4: A simplified graph with a group vertex for the situation shown in Figure 1

## References

Douglas E. Appelt. 1985. Planning English referring expressions. *Artificial Intelligence*, 26:1–33.

Mark Cavazza, Fred Charles, and Steven J. Mead. 2002. Character-based interactive storytelling. *IEEE Intelligent Systems*, 17(4):17–24.

Herbert H. Clark. 1973. Space, time, semantics, and the child. In T. E. Moore, editor, *Cognitive development and the acquisition of language*, pages 65–110. Academic Press.

Robert Dale and Nicholas Haddock. 1991. Generating referring expressions involving relations. In *Proceedings of the Fifth Conference of the Eu-*

*ropean Chapter of the Association for Computational Linguistics (EACL'91)*, pages 161–166.

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. 1992. Generating referring expressions: Constructing descriptions in a domain of objects and processes. MIT Press, Cambridge.

Peter Heeman and Graem Hirst. 1995. Collaborating referring expressions. *Computational Linguistics*, 21(3):351–382.

Annette Herskovits. 1986. *Language and Spatial cognition: an interdisciplinary study of the prepositions in English*. Cambridge University Press.

Helmut Horacek. 1997. An algorithm for generating referential descriptions with flexible interfaces. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 206–213.

Emiel Kraemer and Mariët Theune. 2002. Efficient context-sensitive generation of descriptions. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing: Givenness and Newness in Language Processing*. CSLI Publications, Stanford, California.

Emiel Kraemer, Sebastiaan van Erk, and André Verleg. 2003. Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.

Stephen C. Levinson, editor. 2003. *Space in Language and Cognition*. Cambridge University Press.

Hozumi Tanaka, Takenobu Tokunaga, and Yusuke Shinyama. 2004. Animated agents capable of understanding natural language and performing actions. In Helmut Prendinger and Mituru Ishizuka, editors, *Life-Like Characters*, pages 429–444. Springer.

Kristinn R. Thórisson. 1994. Simulated perceptual grouping: An application to human-computer interaction. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 876–881.

Kees van Deemter. 2002. Generating referring expressions: Boolean extensions of the incremental algorithm. *Computational Linguistics*, 28(1):37–52.

Ielka van der Sluis and Emiel Kraemer. 2000. Generating referring expressions in a multimodal context: An empirically oriented approach. Presented at the CLIN meeting 2000, Tilburg.