Generating Collective Spatial References

Albert Gatt (agatt@csd.abdn.ac.uk)
Department of Computing Science, Meston 216, King's College
Aberdeen AB24 3UE United Kingdom

Abstract

Generation of Referring Expressions is concerned with distinguishing descriptions for target referents in a knowledge base. Plural reference introduces novel problems, one of which is the collective/distributive distinction. This paper presents an empirical study of the production of collective spatial references, and an algorithm that determines content for such expressions from spatial data.

Keywords: Natural Language Generation; psycholinguistics; reference; collective predication.

Introduction

Generation of Referring Expressions (GRE) is an integral part of the microplanning stage of Natural Language Generation (NLG). Given a *target referent*, a GRE algorithm selects properties (or predicates) that will distinguish the target from its *distractors* in a knowledge base (KB), assumed to be accessible to both the system and the user/hearer (Dale & Reiter, 1995). Following Dale and Reiter, this process of *content determination* is usually conceived as a search, during which the GRE algorithm finds properties that are true of the target and not true of at least a subset of the distractors. Such properties are incrementally added to the description, and the process continues until the intersection of all the selected properties has only the target in its extension.

Recent work has extended this problem definition to the generation of plural references. Such work has mostly focused on the two interrelated issues of *logical completeness* and *naturalness*. Completeness requires that a GRE algorithm find a distinguishing description whenever one exists. In the case of plurals, this requires dealing not only with the intersection or conjunction of predicates, but also logical disjunction (set union) (van Deemter, 2002). It turns out however, that extending existing algorithms in this way often results in extremely complex descriptions. Taking their cue from Grice's (1975) notion of Brevity, several authors have proposed ways of controlling the search for a distinguishing description in order to avoid this (e.g. Gardent, 2002; Horacek 2004).

One problem that has recieved very little attention in this area is that of *collective predication*. Plural predicates can have a collective reading, and the difference between such readings and distributives often has consequences for the truth conditions of a sentence. For instance, (1) could be (distributively) true of each of the men in question, and collectively true of a group of men, taken as a whole. A collective reading is often forced by the presence of numerals and grouping operators (2), which make reference to a group explicit. The expression *the group of N X* can only be true if there are *N* individuals of type *X* that form a group.

(1) the men in the corner

(2) the group of four women in the pub

Much work in formal semantics has focused on giving an account of the collective/distributive distinction (e.g. Scha, 1983; Landman, 1989; Landman, 1996). This paper, on the other hand, will be concerned with the automatic generation of collective references to sets or groups, focusing on *spatial* domains. There are three issues to be faced:

P1 **The grounding problem**: When is it appropriate to refer to a set in the domain as a group?

P2 **Expressive choice**: When is it correct to use a grouping operator and/or a numeral?

P3 **Descriptive strategy**: When is reference to a group sufficient and when is it necessary to identify it in relation to its distractors? There are two broad possibilities:

1. **absolute strategy**: distinguish a group based solely on its properties
2. **navigational strategy**: explicitly distinguish a group from its distractors

In what follows, we assume that there is a domain of entities which have, among other things, a location property, represented as a vector of coordinates. Here, the grounding problem (P1) is to determine which sets of entities form a perceptual group, warranting the application of a group operator. This also has a bearing on descriptive strategy (P3). If a set of referents form a group, it may be necessary to distinguish it from distractors that are near enough to it. The question of expressive choice is an open empirical question: it's not clear whether the use of numerals and explicit grouping operators depend on factors such as target set size and distractor proximity.

The rest of this paper introduces some related work, then presents an approach to grounding spatial descriptions. An experiment addressing P2 and P3 is described. Finally, we bring together these two strands in a procedure that determines the content of spatial descriptions, which is assumed to constitute the input to a linguistic realisation module.

Related work

The generation of collectives has been addressed by Stone (2000) within a constraint-based framework. The idea is to find, for a given target set R, a *salient cover* for the set, i.e. a set of properties P such that $\bigcup_{p \in P} \|p\| = R$. However, the empirical question of what constitutes a *salient* cover is not addressed. Funakoshi *et al.* (2004) use empirical data to motivate a navigational generation strategy for singular references in spatial domains. Starting from the entire domain, their algorithm generates a sequence of groups containing the intended referent. The sequence is then rendered linguistically. The groups are identified using the perceptual grouping algorithm proposed by Thorisson (1994). This generation strategy can easily be generalised to sets of objects; however, it's not clear that a sequence-of-groups would be the strategy of choice for non-Japanese speakers, or indeed whether it is necessary to always start from the entire domain.

Gorniak & Roy (2004) also address the problem of grounding in a reference resolution task. Their approach was to encode a direct mapping between linguistic devices used by subjects in (singular) object references and features of the visual scenes those devices referred to. Although it was not carried out in a controlled setting, their study bears a resemblance to the one reported below. Nevertheless, the grounding of linguistic expressions in visual scenes is different in a resolution task, since the decision to refer collectively does not devolve on the system.

Dealing with the grounding problem

The symbol grounding problem, as formulated by Harnad (1990), can be paraphrased as follows: Given a symbol system consisting of a finite alphabet and a set of syntactic rules to manipulate the symbols, how is it possible to formulate the meaning of the symbol tokens without reference to other symbols (which makes the exercise circular)? Problem P1 is an incarnation of this in a limited domain. Consider a reference to $\{A1, B4, C5\}$ in Figure 1. The description *the (group of) three lightbulbs* is a misleading reference, because these elements do not constitute a good perceptual group. Perceptually grounded groups are perceived as identifiable clusters, or 'complex entities' composed of multiple individuals. Wertheimer's (1938) principles of perceptual organisation hold that such groups are perceived when their elements are sufficiently close, and sufficiently distant from other domain elements.

Our solution to the grounding problem is to consider only perceptually well-formed groups as candidates for the application of a grouping operator such as *the group of X*. Implicit in this treatment is a semantic view of groups which, following Landman (1989, 1996), treats them as complex individuals formed by a type-raising operation. The operation, which is triggered by the application of a grouping operator, takes a set and returns a group.

Perceptual groups are discovered by the procedure $makeGroups(D)$, a graph-theoretic generalisation of a clustering algorithm proposed by Gatt (2006). The procedure partitions a domain D into groups which satisfy Wertheimer's Principle of Proxmity, interpreted in terms of the Nearest Neighbour Principle (NNP). NNP holds that elements should be grouped together with their nearest neighbours. D is a

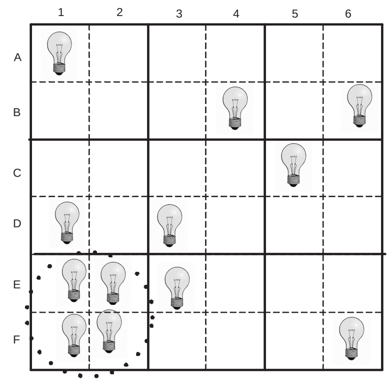

Figure 1: Example domain

set of entities, each of which has a *location* attribute, whose value is a tuple of length n, corresponding to n dimensions. The distance function $\delta(a, b)$, returns the Euclidean distance between any pair of entities. As a first step, a sparse, directed graph $\mathcal{G} = \langle V, E \rangle$ is constructed to represent the nearest neighbour relation, where the set V of vertices represents domain entities, and:

$$\forall e, e' \in V : \left[\langle e, e' \rangle \in E \leftrightarrow \delta(e, e') = \min_{x \in D - \{e\}} \delta(e, x)\right] \tag{3}$$

Each entity $e \in V$ has a single emanating edge, denoted $\mathcal{G}[e]$, leading to its nearest neighbour (hereafter $nn(v, \mathcal{G}[e])$). In its simplest form, grouping could proceed by taking the transitive closure of the nearest neighbour relation: If $nn(e, e')$ and $nn(e', e'')$, then $\{e, e', e''\}$ are clustered together. However, whether two elements satisfying NNP are grouped together also depends on the distance between them. For instance, the nearest neighbour of $F6$ in Figure 1 is $E3$, which is likely to be grouped with $D3$. However, $F6$ is too distant to warrant inclusion in this group.

We resolve this by using data from a study on perceptual grouping (Gatt 2006). 13 participants were shown 4 domains consisting of 13 points constructed by plotting $\langle x, y \rangle$ coordinates. They were asked to partition them into groups according to their intuitions. From this data, we extracted all nearest-neighbour pairs and calculated the proportion of participants who chose to group the pairs together. In all cases, participants observed NNP, grouping entities with their nearest neighbours, if they grouped them at all.

A regression analysis of pairwise distance on response proportions was conducted, to determine the predicted likelihood of acceptance $p(a, b)$ for a grouped pair $\langle a, b \rangle$. This yielded equation (4), which had an optimal fit to the data ($\beta = -.96, p < .001$):

$$p(a, b) = 1.091 + (-.364 \times \delta(a, b)) \tag{4}$$

$makeGroups(D)$ proceeds by looping through the set of objects in D. For each entity (vertex) v, the decision to group v and $\mathcal{G}[v]$ together depends on p. A threshold is set, so that a pair is grouped if $p(v, \mathcal{G}[v]) > .75$. In case one of the two entities is already in some group G, the acceptance rate is calculated between e and the groups's geometric centroid ($cen(G)$):

$$cen(G) = \frac{\sum_{i=1}^{|G|} x_i}{|G|} \quad (5)$$

Evaluation The output of $makeGroups(D)$ was evaluated by calculating an agreement score between the algorithm and each participant for each entity e in each domain[1]. Let G_h be a group proposed by a participant (H), containing e, and let G_a be the group containing e found by the algorithm (A). The score computed by (6) reflects the extent to which human and algorithm agree on the group that e should be placed in.

$$agr(H, A, e) = \frac{|G_h \cap G_a|}{|G_h \cap U_a| + |\overline{G_h \cap U_a}|} \quad (6)$$

Figure 2 displays mean agreement scores on each of the thirteen entities in the four domains. Agreement is over 0.85 in 75% of cases.

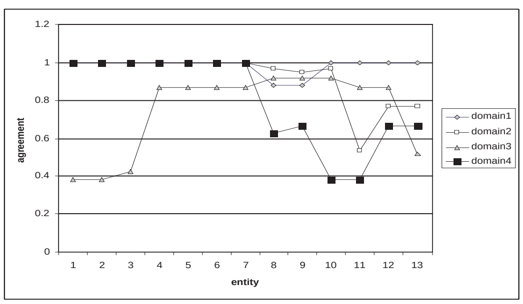

Figure 2: Evaluation of makeGroups(D)

Significant disagreement arose due to reciprocal pairs, that is, cases where $nn(a, b)$ and $nn(b, a)$. In such cases, the algorithm simply groups $\{a, b\}$ together, and no further elements will be added to the group. Humans, on the other hand, merged such a group with neighbouring groups when they were close enough. The solution is to extend the nearest neighbour relation to groups, on which basis we also define a notion of *mergeability*. Both relations are defined over pairs of groups $G_1, G_2 \in makeGroups(D)$, and will be useful in the content determination procedure to be described below.

Definition 1. Group Nearest Neighbour
G_2 is the nearest neighbour of G_1, $nn(G_1, G_2)$, iff:

$$\delta(cen(G_1), cen(G_2)) = \min_c \delta(cen(G_1), cen(c))$$

where $c \in makeGroups(D) - G_1$.

Definition 2. Mergeability
G_1 and G_2 are mergeable, $mergeable(G_1, G_2)$, iff:

$$p(cen(G_1), cen(G_2)) > 0.75$$

[1]See Gatt (2006) for a comparison to Thorisson's (1994) algorithm.

Grounded semantics for group expressions Since the output of $makeGroups(D)$ closely approximates human intuitions, it is used to determine whether a given target set of referents is a good candidate for a collective group description. We first define the notion of a *spatially grounded group*.

Definition 3. Spatial Grounding
$S \subseteq D$ is spatially grounded iff:

$$\exists G \in makeGroups(D) : S \subseteq G$$

Following Landman (1989), grouping operators such as *the cluster* are viewed as linguistic counterparts of the semantic operation mapping pluralities, which denote sets, to groups, which are complex individuals. This is restricted by the grounding condition.

Investigating descriptive strategies

This section describes some results from an experiment that addressed the issues introduced under P2 and P3 in the Introduction.

Design and procedure

The experiment involved a task which required participants to refer to a target set in a 2D domain where the only feature that distinguished the elements of the set was their location.

Materials and design A domain consisted of a 3×3 matrix, as shown in Figure 1. This created 9 possible locations. Each space was further subdivided into 4 cells (the 'object cells') of equal size. Domain objects were selected from the colour version of the Snodgrass/Vanderwart normed picture set (Rossion & Pourtois, 2004). A trial consisted of a target set in one of the 9 locations, marked by a faint dotted red circle, surrounded by 8 distractors of identical type. The grid structure was not visible to participants. Two factors were manipulated in the experiment:

1. **Cardinality** Target sets contained 1 (*C1*), 4 (*C4*), or 8 (*C8*) objects.

2. **Distractors** Distractors were *close* to or *distant* from the target set. In the *close* condition, distractors in locations adjacent to the target location were placed in object cells bordering the target location, as in Figure 1. In the *distant* condition, at least one object cell separated a distractor from the border of the target location.

Three sets of 18 trials each were constructed, each of which contained 6 trials in each level of **cardinality**, half of which were in the *close* and half in the *distant* **distractor** conditions. Trial sets satisfied two constraints: (a) there were two trials with targets in each of the nine locations; (b) every **cardinality** $\times$ **distractor** combination was in a different location in each version.

We expected **distractors** to strongly impact descriptive strategy, with more navigational responses overall when distractors were *close*. However, since large groups are potentially more salient, **cardinality** should also impact strategy, with fewer navigational descriptions in $C8$, and more in $C1$. If set size increases salience, grouping operators should be predominant in $C8$ relative to $C4$.

Table 1: Proportions (%) of response types

Cardinality	Close				Distant			
	abs	nav	grp	num	abs	nav	grp	num
$C1$	33	55	NA	NA	81.7	6	NA	NA
$C4$	80	16.7	60	15	100	0	40	35
$C8$	76.2	22.2	57.1	58.7	96.8	3.2	57.1	58.7

Participants and method 20 (self-rated) native or fluent speakers of English carried out the experiment remotely on the internet[2]. They were asked to imagine playing a game which required them to identify a set of objects for a partner. Participants were randomly assigned to one of the three versions; trials were randomised. A trial consisted of a domain, and a question (*Which objects are in the dotted red circle?*), with a text box in which participants typed responses.

Results and discussion

The results consist of a corpus of ca. 360 descriptions, approximately equally divided into references to 1, 4 and 8 objects. Descriptions were classified according to descriptive strategy and collective expression type as follows:

1. **descriptive strategy**
   (a) *navigational*: any description to the target set that involved reference to the distractors. E.g. *the cluster of 7 balloons: there's one balloon above and one balloon below the cluster*
   (b) *absolute*: any description that involved reference solely to the target set. E.g. *the four cups closest to the top right corner*
   (c) *other*
2. **collective expression** ($C4$ and $C8$ conditions only)
   (a) *+/- grouping operator*: any description that contained a group expression and/or a reference to the shape of the target set. E.g. *the 8 bulbs in the conspicuous group on the top*
   (b) *+/- numeral*: any description that contained an explicit mention of the number of objects in the target set.

Results Table 1 displays proportions of responses by condition. We report results of a by-items analysis (χ^2_1) on response frequencies. By participants, we use a Friedman analysis (χ^2_2) on proportions of response types, and signed rank tests (Z) for pairwise comparisons.

The expectations regarding the effect of **cardinality** on grouping and numerals were not borne out: Grouping operators were equally common in $C4$ and $C8$ conditions ($\chi^2_1 = 1.26, ns$, $\chi^2_2 = 1.47, ns$). There was a greater tendency to explicitly use numerals in the $C8$ condition, reliable both by items ($\chi^2_1 = 28.66, p < .001$) and participants ($\chi^2_2 = 13.00, p < .001$). The use of numerals and grouping expressions was not dependent on the closeness of **distractors** (*close:* $\chi^2_1 = 2.40, ns$, $\chi^2_2 = 1, ns$, *distant:* $\chi^2_1 = 2.35, p > .1$, $\chi^2_2 = .8, ns$).

The most common descriptive strategy overall was the *absolute* strategy. However, there was a significant effect of **cardinality** ($\chi^2_1 = 32.75, p < .001$), with a greater proportion of *navigational* descriptions in the $C1$ condition. The difference between **cardinality** levels was significant overall in proportions of both *absolute* ($\chi^2_2 = 6.58, p = .03$) and *navigational* responses ($\chi^2_2 = 16.11, p < .001$). Pairwise comparisons within the *close* condition showed that navigational response proportions were reliably different in the *close* condition between $C1$ and $C4$ ($Z = 3.012, p = .003$), and $C1$ and $C8$ ($Z = 1.949, p = .05$), but not $C4$ and $C8$. Within the *distant* condition, proportions of *navigational* responses were only reliably different between $C1$ and $C8$ ($Z = 2.23, p = .03$), with the $C1 - C4$ contrast barely reaching significance ($Z = 1.79, p = .07$).

Descriptive strategy was strongly dependent on the closeness of **distractors**, with significantly more *absolute* responses in the *distant* condition ($\chi^2_1 = 47.07, p < .001$). By participants, the proportions of both *absolute* responses differed significantly across the two conditions (absolute: $\chi^2_2 = 16, p < .001$, navigational: $\chi^2_2 = 11.27, p = .001$). The difference in response proportions between *close* and *distant* conditions was significant at all three levels of **cardinality**: $C1$: $Z = 3.08, p = .002$; $C4$: $Z = 1.97, p = .05$; and $C8$: $Z = 2.68, p = .007$. Proportions of *navigational* responses displayed a similar pattern, except in the $C4$ condition, where the difference failed to reach significance ($C1: Z = 3.04, p = .002$, $C4: Z = 1.82, p = .07$, $C8: Z = 2.04, p = .002$).

General discussion The use of grouping operators and numerals was predominant irrespective of set size or of distractor proximity. The fact that larger sets tend to be described more often with reference to their size is probably due to the greater precision afforded by numerals when talking about relatively large quantities, thus ensuring an unambiguous description.

One possible reason for the overall frequency of the *absolute* strategy is that participants expected location and cardinality information to be sufficient for identification of large, and therefore salient, sets. For example, one participant wrote *the group of 8 bottles in the bottom right of the screen*, where distractors were close to the target. The fact that such responses were less prevalent in $C1$ supports this claim. However, strategy was highly dependent on distractor closeness. Indeed, $C1$ only differed reliably from $C8$ in the *distant* condition. In some of these cases, an isolated referent could have been perceived as mergeable with a nearby group, motivating a *navigational* strategy. Singletons are also less salient than

[2] Precautions were taken to disallow participants from repeating the experiment.

large groups. In this connection, it is interesting to note that there were descriptions in the $C4$ and $C8$ conditions which made explicit reference to group salience (7)

(7) the 8 balloons forming a conspicuous group

On closer examination, two main types of *navigational* responses were distinguished: (a) the *subset* strategy (8) where the target is identified within a containing group; (b) a *complement* strategy (9), identifying the target and explicitly describing surrounding distractors. Of these, (a) was the most frequent, comprising 67% of all navigational responses.

(8) the group of lamps on the right middle . . . take the 4 lamps of the bottom . . . AND the corresponding 4 above them

(9) the eight ones clustered around the middle . . . this leaves one object out . . . left of the middle

While the *subset* strategy has the flavour of a 'sequence of groups', as proposed by Funakoshi *et al.* (2004), it was usually limited to a reference to the immediately containing set, followed by a reference to the target.

Content determination

This section brings together the results reported, describing a content-determination strategy for collective descriptions of spatial groups. Our results suggest that descriptive strategy depends on whether there are distractors close to the target set, which occurs in two cases: (a) the target is a proper subset of a grounded group; and/or (b) the nearest neighbour of the target set's containing group is mergeable with the group. The core of the content-determination procedure, shown in Algorithm 1 builds a *description set* for the target.

Definition 4. [Description set] A description set $dSet$ for a target referent set R is a set of tuples $\langle R', G, N\rangle$ where:

1. G and N are grounded in D
2. either $\big(nn(G,N) \wedge mergeable(G,N)\big)$ or $N = \emptyset$
3. $\big(R' \subseteq R\big) \wedge \big(R' \cap G \neq \emptyset\big)$
4. $\bigcup_{R' \in dSet} = R$

In these tuples, R' corresponds to (a subset of) the target satisfying the groundedness condition (Def 3). This is trivially satisfied in case $|R'| = 1$. The set G is the spatially grounded group containing R'. If $G - R' \neq \emptyset$, then the elements of R' need to be distinguished from $G - R'$, calling for a navigational strategy. Similarly, if the containing group has a mergeable nearest neighbour N, R' needs to be distinguished from this too.

The procedure $descriptionSet(D, R)$ iterates through the groups returned by $makeGroups(D)$. If a group and the target R have a nonempty intersection (1.3), the desciption set is updated with a new tuple, whose third element N is either a mergeable nearest neighbour of the containing group G, or empty (1.4–1.9). Target elements thus accounted for are removed (1.11). Since $makeGroups(D)$ partitions the domain, for any element in the target set, there will be one and only one group that contains that element.

Algorithm 1 descriptionSet(D, R)

1: $dSet \leftarrow \emptyset$
2: **for all** $G \in makeGroups(D)$ **do**
3:   **if** $R \cap G \neq \emptyset$ **then**
4:     $N \leftarrow nearestNeighbour(G)$
5:     **if** $mergeable(N, g)$ **then**
6:       $dSet \leftarrow dSet \cup \langle R \cap G, G, N\rangle$
7:     **else**
8:       $dSet \leftarrow dSet \cup \langle R \cap G, G, \emptyset\rangle$
9:     **end if**
10:   **end if**
11:   $R \leftarrow R - G$
12:   **if** $R = \emptyset$ **then**
13:     **return** $dSet$
14:   **end if**
15: **end for**

A description set constitutes the input to a procedure that fleshes out its content. Each tuple in the set is mapped to a description that is maximally of the form *the group of R' in G excluding N*, where G, R', N are described in the following way.

1. If $G - R'$ is nonempty, return a description of the form $grp(R')$ *in* $grp(G)$. Otherwise return $grp(R')$.
2. If N is nonempty, return $grp(N)$
3. $grp(S)$, for any S is constructed as follows:
   (a) If S contains 4 or more elements, $grp(S)$ consists of a group operator, a numeral quantifier, and the location of the set within its smallest containing set.
   (b) If S contains between 2 and 4 elements, $grp(S)$ consists of a numeral and location only.
   (c) If S is a singleton, $grp(S)$ contains location information only.

In line with the experimental findings, numerals are always used except in the case of singletons, while group operators are omitted if the set has fewer than four elements. Navigational strategies are used when G and/or N are nonempty. In case R' is contained within a larger group, the *subset* strategy is used, describing R' in terms of its location in its smallest containing set, that is G. A mergeable nearest neighbour, by contrast, is excluded by relating G to it in what is essentially the *complement* strategy. Note that both G and N are described in absolute terms, since their smallest containing group is the domain itself.

Examples

The following examples are given with reference to Figure 1. We will first consider the case where the target set is $\{E1, E2, F1, F2\}$. The procedure $makeGroups(D)$ returns, among others, the following groups: $1 : \{E1, E2, F1, F2, D1\}$, $2 : \{D3, E3\}$. Group 2 is not merged with Group 1, because $D3$ and $E3$ are a reciprocal pair. The call to $descriptionSets(D, R)$ identifies Group 1 as a spatially grounded group which contains all the referents. As a result, no more referents remain to be distinguished at line (1.11). The nearest neighbour of Group 1, Group 2, is mergeable. Hence, the description set returned contains the tuple $\langle\{E1, E2, F1, F2\}, \{E1, E2, F1, F2, D1\}, \{D3, E3\}\rangle$. Since the target set is contained in Group 1, this part of the description is fleshed out by a group description of the target in the containing group. This part of the description has the form $grp(R')$ *in* $grp(G)$, where $grp(R')$ contains a numeral, a group operator, and the target's location within G, while $grp(G)$ is described in absolute terms. Since N is nonempty, it is also described using numeral and location information. The end result would be realised as shown in (10).

(10) **the group of four bulbs** in **the group of five at the bottom left** excluding **the two objects towards the middle**

Although there is considerable variation in the corpus, this output is comparable to several human descriptions in similar domains, such as (11), which first identifies the containing set (*9 bulbs*) via its location, excluding the distractor.

(11) The ... 8 of 9 bulbs closest to the bottom leftmost corner of the frame ... A ninth bulb in this cluster, the one lying closer to the central part ...

Consider next a reference to $\{B4, B6, F6\}$. In this case, the two relevant grounded groups are $\{B4, C5, B6\}$ and $\{F6\}$. None of these two groups have a mergeable nearest neighbour. The description set contains two tuples: $\langle\{B4, B6\}, \{B4, B5, B6\}, \emptyset\rangle$ and $\langle\{F6\}, \{F6\}, \emptyset\rangle$. The first of these results in a reference to $\{B4, B6\}$ in terms of its location within the containing set. For the second, since $G - R' = \emptyset$ and there is no mergeable nearest neighbour, the singleton is described in absolute terms.

Conclusions

This paper has discussed the generation of collective spatial descriptions, focusing on the problems of grounding the semantics of group operators in perceptual data, and finding adequate descriptive strategies. The approach was informed by two empirical studies on perceptual grouping and collective spatial reference.

Current work is exploiting the corpus further. While the present paper focused on *when* and *how* to refer using group operators and numerals, complications arise with *lexical choice* and realisation, where the corpus evinces considerable individual variation. The data on descriptive strategies is also being studied more closely, with a view to evaluating the output of the algorithm more systematically against the human output.

References

Dale, R., & Reiter, E. (1995). Computational interpretation of the Gricean maxims in the Generation of Referring Expressions. *Cognitive Science*, *19*, 233–263.

Funakoshi, K., Watanabe, S., Kuriyama, N., & Tokunaga, T. (2004). Generating referring expressions using perceptual groups. *Proceedings of the 3rd International Conference in Natural Language Generation*

Gardent, C. (2002). Generating minimal definite descriptions. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.

Gatt, A. (2006). Structuring knowledge for reference generation: A clustering algorithm. *Proceedings of the 11th Conference of the European Chapter of the ACL*.

Gorniak, P., & Roy, D. (2004). Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, *21*, 429–470.

Grice, P. (1975). Logic and conversation. In: P. Cole and J. Morgan (Eds.), *Syntax and Semantics: Speech Acts*. New York: Academic Press.

Harnad, S. (1990). The symbol grounding problem. *Physica*, *42*, 335–346.

Horacek, H. (2004). On referring to sets of objects naturally. *Proceedings of the 3rd International Conference in Natural Language Generation*

Landman, F. (1989). Groups I. *Linguistics and Philosophy*, *12*, 559–605.

Landman, F. (1996). Plurality. In: S. Lappin (Ed.), *Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Reiter, E., & Sripada, S. (2002). Human variation and lexical choice. *Computational Linguistics*, *28*, 545–553.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object databank: The role of surface detail in basic level object recognition. *Perception*, *33*, 216–236.

Scha, R.J.H. (1980). Distributive, collective, and cumulative quantification. In: J. Groenendijk, M. Stokhof, and T. Janssen (Eds.), *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum.

Stone, M. (2000). On identifying sets. *Proceedings of the 1st International Conference in Natural Language Generation*

Thorisson, K. (1994). Simulated perceptual grouping: An application to Human-Computer Interaction. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*.

van Deemter, K. (2002). Generating referring expressions: Boolean extensions of the Incremental Algorithm. *Computational Linguistics*, *28*, 37–52.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In: W. Ellis (Ed.), *A Source Book of Gestalt Psychology*. London: Routledge and Kegan Paul.

Acknowledgements

Thanks to Kees van Deemter, Ehud Reiter, Graeme Ritchie, Judith Masthoff and three anonymous reviewers for helpful comments. This work forms part of the TUNA Project (http://www.csd.abdn.ac.uk/ agatt/tuna), supported by EPSRC Grant No. GR/S13330/01.