Incremental generation of plural descriptions: Similarity and partitioning

Albert Gatt and **Kees van Deemter**
Department of Computing Science
University of Aberdeen
{agatt,kvdeemte}@csd.abdn.ac.uk

Abstract

Approaches to plural reference generation emphasise simplicity and brevity, but often lack empirical backing. This paper describes a corpus-based study of plural descriptions, and proposes a psycholinguistically-motivated algorithm for plural reference generation. The descriptive strategy is based on partitioning. An exhaustive evaluation shows that the output closely matches human data.

1 Introduction

Content Determination for the Generation of Referring Expressions (GRE) starts from a Knowledge Base (KB) consisting of a set of entities U and a set of properties $\mathbb{P}$ represented as attribute-value pairs, and searches for a description $D \subseteq \mathbb{P}$ which distinguishes a referent $r \in U$ from its distractors. For example, the KB in Table 1 represents 8 entities in a $2D$ visual domain, each with 6 attributes, including their location, represented as a combination of horizontal (X) and vertical (Y) numerical coordinates. To refer to an entity an algorithm searches through values of the different attributes.

GRE has been dominated by Dale and Reiter's (1995) Incremental Algorithm (IA), one version of which, generalised to deal with *non-disjunctive* plural references[1] is shown in Algorithm 1 (van Deemter, 2002). After initialising the description D and the distractor set C [1.1–1.2], IA$_{plur}$ traverses an ordered list of properties, called the *preference order* ($\mathcal{PO}$) [1.3], which reflects general or domain-specific preferences for attributes. For instance, with the $\mathcal{PO}$ in the top row of the Table, the algorithm first considers values of TYPE, then COLOUR, and so on, adding a property to D if it is true of the intended referents R, and excludes some

---

[1] *Non-disjunctive* descriptions, such as *the large red chairs*, are logically a conjunction of literals. In *disjunctive* descriptions such as *the chair and the table*, the *and* represents set union (of things which are chairs or tables).

	TYPE	COLOUR	ORIENTATION	SIZE	X	Y
e_1	desk	red	back	small	3	1
e_2	sofa	blue	back	small	5	2
e_3	desk	red	back	large	1	1
e_4	desk	red	front	large	2	3
e_5	desk	blue	right	large	2	4
e_6	sofa	red	back	large	4	1
e_7	sofa	red	front	large	3	3
e_8	sofa	blue	back	large	3	2

Table 1: A visual domain

distractors [1.4]. The description and the distractor set C are updated accordingly [1.5–1.6], and the description returned if it is distinguishing [1.7].

Compared to some predecessors which emphasised brevity (Dale, 1989), the IA is highly efficient, because the use of the $\mathcal{PO}$ avoids exhaustive combinatorial search, potentially overspecifying the description. Overspecification and the use of a $\mathcal{PO}$ have been justified on psycholinguistic grounds. Speakers overspecify their descriptions because they begin their formulation without exhaustively scanning a domain (Pechmann, 1989), terminating the process as soon as a referent is distinguished (Belke and Meyer, 2002). They prioritise the basic-level category (TYPE) of an object, and salient, absolute properties like COLOUR

Algorithm 1 IA$_{plur}(R, U, \text{PO})$

1: $D \leftarrow \emptyset$
2: $C \leftarrow U - R$
3: **for** $\langle \text{A} : v \rangle \in \mathcal{PO}$ **do**
4:     **if** $R \subseteq [\![\langle \text{A} : v \rangle]\!] \wedge [\![\langle \text{A} : v \rangle]\!] - C \neq \emptyset$ **then**
5:         $D \leftarrow D \cup \{\langle \text{A} : v \rangle\}$
6:         $C \leftarrow C \cap [\![\langle \text{A} : v \rangle]\!]$
7:         **if** $[\![D]\!] = R$ **then**
8:             **return** D
9:         **end if**
10:     **end if**
11: **end for**
12: **return** D

(Pechmann, 1989; Eikmeyer and Ahlsèn, 1996), as well as locative properties in the vertical dimension (Arts, 2004). Relative attributes like SIZE are avoided unless absolutely required for identification (Belke and Meyer, 2002). This evidence suggests speakers conceptualise referents as *gestalts* (Pechmann, 1989) whose core is the basic-level TYPE (Murphy, 2002) and some other salient attributes like COLOUR. Note that the IA does not fully mirror these human tendencies, since it only includes preferred attributes in a description if they remove some distractors, whereas psycholinguistic research suggests that people include them irrespective of contrastiveness (cf. van der Sluis and Krahmer, 2005).

More recent research on plural GRE has de-emphasised these issues, especially in case of *disjunctive* plural reference. The first concrete proposal in this area, IA_{bool} (van Deemter, 2002), first tries to find a non-disjunctive description using Algorithm 1. Failing this, it searches through disjunctions of properties of increasing length, generating a description in Conjunctive Normal Form (CNF). For example, calling the algorithm with $R = \{e_1, e_2\}$ would result in a non-disjunctive description, since both referents can be distinguished using ⟨SIZE : *small*⟩. However, a conjunction wouldn't suffice to distinguish $R = \{e_1, e_8\}$, and IA_{bool} would consider combinations such as ⟨TYPE : *desk*⟩ ∨ ⟨COLOUR : *blue*⟩. This generalised algorithm has three consequences:

1. **Efficiency**: Searching through disjunctive combinations results in a combinatorial explosion (van Deemter, 2002).

2. **Gestalts and content**: The notion of a 'preferred attribute' is obscured, since it is difficult to apply the same reasoning that motivated the $\mathcal{PO}$ in the IA to combinations like (COLOUR ∨ SIZE).

3. **Form**: Descriptions can become logically very complex (Gardent, 2002; Horacek, 2004).

Some proposals to deal with (3) include Gardent's (2002) non-incremental, constraint-based algorithm to generate the briefest available description of a set. An alternative, by Horacek (2004), combines best-first search with optimisation to reduce logical complexity. Neither approach benefits from empirical grounding, and both leave open the

question of whether previous psycholinguistic research on singular reference is at all applicable to the plural disjunctive case.

This paper starts with an empirical analysis of plural descriptions using a semantically transparent corpus of descriptions elicited in well-defined domains, of which Table 1 is an example. Based on the data analysis, we propose and evaluate an efficient algorithm for the generation of references to arbitrary sets. Our starting point is the assumption that plurals, like singulars, evince preferences for certain attributes. Based on previous work in Gestalt perception (Wertheimer, 1938; Rock, 1983), we propose an extension of Pechmann's Gestalts Principle, whereby plural descriptions are preferred if (a) *they maximise the similarity of their referents*, using the same attributes to describe them as far as possible; (b) prioritise salient ('preferred') attributes which are central to the conceptual representation of an object. We address (3) above by investigating the logical form of plurals in the corpus. One strong determinant of descriptive form is the basic-level category of objects. For example, to refer to $\{e_1, e_2\}$ in the Table, an author has at least the following options:

(1) (a) the small desk and sofa
     (b) the small red desk and the small blue sofa
     (c) the small desk and the small blue sofa
     (d) the small objects

We refer to (1a) as an *aggregated disjunctive description*, in that the property *small* has wide scope scope over the coordinate NP *desk and sofa* (which is logically a disjunction). By contrast, (1b,c) are *non-aggregated* and overspecified because they contain COLOUR when SIZE alone suffices. The most economical description is (1d), which is *non-disjunctive*. This is possible because it contains a superordinate TYPE (*object*). Since basic-level categorisation is preferred on independent grounds (Rosch et al., 1976), we expect (1a–c) to be more frequent. Note that (1b,c) represent a *partition* of R and describe each element separately. In (1b), there is considerable redundancy in including COLOUR twice. The potential benefit of this is that the elements of the partition are described in a *parallel* fashion, using exactly the same attributes (SIZE and COLOUR). This is not the case in (1c), which is *non-parallel*. By hypothesis, parallelism adds to the perceptual cohesion of the set. Given the psycholin-

```
<DESCRIPTION num='pl'>
<DESCRIPTION num='singular'>
<ATTRIBUTE name='size' value='small'>small</ATTRIBUTE>
<ATTRIBUTE name='colour' value='red'>red</ATTRIBUTE>
<ATTRIBUTE name='type' value='desk'>desk</ATTRIBUTE>
</DESCRIPTION>
and
<DESCRIPTION num='sg'>
<ATTRIBUTE name='size' value='small'>small</ATTRIBUTE>
<ATTRIBUTE name='colour' value='blue'>blue</ATTRIBUTE>
<ATTRIBUTE name='type' value='sofa'>sofa</ATTRIBUTE>
</DESCRIPTION>
</DESCRIPTION>
```

$$(\langle \text{SIZE} : small \rangle \wedge \langle \text{COLOUR} : red \rangle \wedge \langle \text{TYPE} : desk \rangle)$$
$$\vee$$
$$(\langle \text{SIZE} : small \rangle \wedge \langle \text{COLOUR} : blue \rangle \wedge \langle \text{TYPE} : sofa \rangle)$$

Figure 1: Corpus annotation examples

guistic evidence, the hypothesised tendency to emphasise similarity may also be somewhat dependent on the attributes involved, so that COLOUR should be more likely to be redundantly propagated across disjuncts than a relatively dispreferred attribute like SIZE.

2 The data

The data for our study is a subset of a corpus of 900 references to furniture and household items, collected via a controlled experiment involving 45 participants. In addition to their TYPE, objects in the domains have COLOUR, ORIENTATION and SIZE. For each subset of these three attributes, there was an equal number of domains in which the minimally distinguishing description (MD) consisted of values of that subset. For example, Table 1 represents a domain in which the intended referents, $\{e_1, e_2\}$, can be minimally distinguished using only SIZE[2]. Thus, overspecified usage of attributes can be identified in authors' descriptions. Domain objects were randomly placed in a 3 (row) $\times$ 5 (column) grid, represented by X and Y in Table 1. These are relevant for a subset of descriptions which contain locative expressions.

Corpus descriptions are paired with an explicit XML domain representation, and annotated with semantic markup which makes clear which attributes a description contains. The markup also enables the compositional derivation of a logical form[3]. For example, the XML representation of (1b) is shown in Figure 1, which also displays the LF derived from it.

Descriptions in the corpus were elicited in 7 domains with one referent, and 13 domains with 2 referents. Plural domains represented levels of a *Value Similarity* factor. In 7 *Value-Similar* (VS)

---

[2] TYPE was not included in the calculation of MD.

[3] For details of corpus design and annotation, we refer to (van der Sluis et al., 2006).

	VS		VDS	
	+Disj	−Disj	+Disj	−Disj
+aggr	20.2	15.5	2.4	3.7
−aggr	64.3	−	93.9	−
% overall	84.5	15.5	96.3	3.7

Table 2: % disjunctive and non-disjunctive plurals

domains, referents were identifiable using identical values of the minimally distinguishing attributes. In the remaining 6 *Value-Dissimilar* (VDS) domains, the minimally distinguishing values were different. Table 1 represents a VS domain, where $\{e_1, e_2\}$ can be minimally distinguished using the same value of SIZE (*small*). Thus, MD in VS was a logical conjunction. In VDS, it was a disjunction since, if two referents could be minimally distinguished by different values v and v' of an attribute A, then MD had the form $\langle A : v \rangle \vee \langle A : v' \rangle$. However, even in VS, referents had different basic-level types. Thus, an author faced with a domain like Table 1 had at least the descriptive options in (1a–d).

Our analysis will focus on a stratified random sample of 180 plural descriptions, referred to as PL_1, generated by taking 4 descriptions from each author (2 each from VS and VDS conditions). We also use the singular data (SG; $N = 315$). The remaining plural descriptions (PL_2; $N = 405$) are used for evaluation.

3 The logical form of plurals

Descriptions in PL_1 were first classified according to whether they were non-disjunctive (cf. 1c) or disjunctive (1a,b). The latter were further classified into aggregated (1a) and non-aggregated (1b). Table 2 displays the percentage of descriptions in each of the four categories, within each level of Value Similarity. Disjunctive descriptions were a majority in either condition, and most of these were non-aggregated. As noted in relation to (1b), these descriptions correspond to partitions of the set of referents.

Since referents in VS had identical properties except for TYPE values, the most likely reason for the majority of disjunctives in VS is that people's descriptions represented a partition of a set of referents induced by the basic-level category of the objects. This is strengthened by the finding that the likelihood of a description being disjunctive or non-disjunctive did not differ as a function of Value Similarity ($\chi^2 = 2.56$, $p > .1$). A χ^2 test on overall frequencies of aggregated versus non-aggregated

	Parallel	Non-Parallel	χ^2 ($p \leq .001$)
overspec.	24.6	75.4	92.467
underspec.	5.3	94.7	42.217
well-spec.	11	89	26

Table 3: Parallelism: % per description type

disjunctives showed that the non-aggregated descriptions ('true' partitions) were a significant majority ($\chi^2 = 83.63$, $p < .001$). However, the greater frequency of aggregation in VS compared to VDS turned out to be significant ($\chi^2 = 15.498$, $p < .001$). Note that the predominance of non-aggregated descriptions in VS implies that properties are repeated in two disjuncts (resp. coordinate NPs), suggesting that a certain kind of redundancy is not problematic (*contra*, for example, Gardent, 2002).

3.1 Conceptual gestalts and similarity

Allowing for the independent motivation for set partitioning based on TYPE values, we suggested in §1 that *parallel* descriptions such as (1b) may be more likely than *non-parallel* ones (1c), since the latter does not use the same properties to describe the two referents. Similarity, however, should also interact with attribute preferences.

For this part of the analysis, we focus exclusively on the disjunctive descriptions in PL_1 ($N = 150$) in both VS and VDS. The descriptions were categorised according to whether they had *parallel* or *non-parallel* semantic structure. Evidence for Similarity interacting with attribute preferences is strongest if it is found in those cases where an attribute is overspecified (i.e. used when not required for a distinguishing description). In those cases where corpus descriptions do not contain locative expressions (the X and/or Y attributes), such an overspecified usage is straightforwardly identified based on the MD of a domain. This is less straightforward in the case of locatives, since the position of objects was randomly determined in each domain. Therefore, we divided descriptions into three classes: A description is *underspecified* if it does not include a locative expression and omits some MD attributes. A description is *overspecified* if either (a) it does not omit any MD attributes, but includes locatives and/or non-required visual attributes; or (b) it omits some MD attributes, but includes both a locative expression and other, non-required attributes. A description is *well-specified* otherwise.

Proportions of Parallel and Non-Parallel descrip-

	Actual		Predicted
	$p(\text{A}, \text{SG})$	$p(\text{A}, \text{PPS})$	$p(\text{A}, \text{PPS})$
COLOUR	.680	.835	.61
SIZE	.290	.359	.28
ORIENTATION	.280	.269	.26
X-DIMENSION	.440	.517	.52
Y-DIMENSION	.630	.647	.65

Table 4: Actual and predicted usage probabilities

tions for each of the three classes are are shown in Table 3. In all three description types, there is an overwhelming majority of Parallel descriptions, confirmed by a χ^2 analysis. The difference in proportions of description types did not differ between VS and VDS ($\chi^2 < 1$, $p > .8$), suggesting that the tendency to redundantly repeat attributes, avoiding aggregation, is independent of whether elements of a set can be minimally distinguished using identical values.

Our second prediction was that the likelihood with which an attribute is used in a parallel structure is a function of its overall 'preference'. Thus, we expect attributes such as COLOUR to feature more than once (perhaps redundantly) in a parallel description to a greater extent than SIZE. To test this, we used the SG sample, estimating the overall probability of occurrence of a given attribute in a singular description (denoted $p(\text{A}, \text{SG})$), and using this in a non-linear regression model to predict the likelihood of usage of an attribute in a plural partitioned description with parallel semantic structure (denoted $p(\text{A}, \text{PPS})$). The data was fitted to a regression equation of the form $p(\text{A}, \text{PPS}) = k \times p(\text{A}, \text{SG})^S$. The resulting equation, shown in (2), had a near-perfect fit to the data ($R^2 = .910$)[4]. This is confirmed by comparing actual probability of occurrence in the second column of Table 4, to the predicted probabilities in the third column, which are estimated from singular probabilities using (2).

$$p(\text{A}, \text{PPS}) = .713 \, p(\text{A}, \text{SG})^{.912} \qquad (2)$$

Note that the probabilities in the Table confirm previous psycholinguistic findings. To the extent that probability of occurrence reflects salience and/or conceptual importance, an order over the three attributes COLOUR, SIZE and ORIENTATION can be deduced (C>>O>>S), which is compatible with the findings of Pechmann (1989), Belke and Meyer (2002) and others. The locative attributes are

---

[4]A similar analysis using linear regression gave essentially the same results.

also ordered (Y>>X), confirming the findings of Arts (2004) that vertical location is preferred. Orderings deducible from the SG data in turn are excellent predictors of the likelihood of 'propagating' an attribute across disjuncts in a plural description, something which is likely even if an attribute is redundant, modulo the centrality or salience of the attribute in the mental gestalt corresponding to the set. Together with the earlier findings on logical form, the data evinces a dual strategy whereby (a) sets are partitioned based on basic-level conceptual category; (b) elements of the partitions are described using the same attributes to if these attributes are easily perceived and conceptualised. Thus, of the descriptions in (1) above, it is (1b) that is the norm among authors.

4 Content determination by partitioning

In this section we describe IA_{part}, a partitioning-based content determination algorithm. Though presented as a version of the IA, the basic strategy is generalisable beyond it. For our purposes, the assumption of a preference order will be maintained. IA_{part} is distinguished from the original IA and IA_{bool} (cf. §1) in two respects: (a) It induces partitions opportunistically based on KB information, and this is is reflected in the way descriptions are represented. (b) The criteria whereby a property is added to a description include a consideration of the overall salience or preference of an attribute, and its contribution to the conceptual cohesiveness of the description. Throughout the following discussion, we maintain a running example from Table 1, in which $R = \{e_1, e_2, e_5\}$.

4.1 Partitioned descriptions

IA_{part} generates a *partitioned description* (D_{part}) of a set R, corresponding to a formula in Disjunctive Normal Form. D_{part} is a set of *Description Fragments* (DFs). A DF is a triple $\langle R_{\text{DF}}, T_{\text{DF}}, M_{\text{DF}} \rangle$, where $R_{\text{DF}} \subseteq R$, T_{DF} is a value of TYPE, and M_{DF} is a possibly empty set of other properties. DFs refer to disjoint subsets of R. As the representation suggests, TYPE is given a special status. IA_{part} starts by selecting the basic-level values of TYPE, partitioning R and creating a DF for each element of the partition on this basis. In our example, the selection of TYPE results in two DFs, with M_{DF} initialised to empty:

(3) DF_1 $\langle \{e_1, e_5\}, \langle \text{TYPE} : desk \rangle, \emptyset \rangle$
    DF_2 $\langle \{e_2\}, \langle \text{TYPE} : sofa \rangle, \emptyset \rangle$

Algorithm 2 $updateDescription(\langle \text{A} : V \rangle, R')$

for $\langle R_{\text{DF}}, T_{\text{DF}}, M_{\text{DF}} \rangle \in D_{part}$ **do**
  if $R' = \emptyset$ **then**
    return
  else if $R_{\text{DF}} \subseteq R'$ **then**
    $M_{\text{DF}} \leftarrow M_{\text{DF}} \cup \{\langle \text{A} : v \rangle\}$
    $R' \leftarrow R' - R_{\text{DF}}$
  else if $R_{\text{DF}} \cap R' \neq \emptyset$ **then**
    $R_{new} \leftarrow R_{\text{DF}} \cap R'$
    $DF_{new} \leftarrow \langle R_{new}, T_{\text{DF}}, M_{\text{DF}} \cup \{\langle \text{A} : v \rangle\} \rangle$
    $D_{part} \leftarrow D_{part} \cup \{DF_{new}\}$
    $R_{\text{DF}} \leftarrow R_{\text{DF}} - R_{new}$
    $R' \leftarrow R' - R_{new}$
  end if
end for
if $\text{A} = \text{TYPE}$ **then**
  $D_{part} \leftarrow D_{part} \cup \{\langle R', \langle \text{A} : v \rangle, \emptyset \rangle\}$
else
  $D_{part} \leftarrow D_{part} \cup \{\langle R', \bot, \{\langle \text{A} : v \rangle\} \rangle\}$
end if

Although neither DF is distinguishing, R_{DF} indicates which referents a fragment is *intended* to identify. In this way, the algorithm incorporates a 'divide-and-conquer' strategy, splitting up the referential intention into 'sub-intentions' to refer to elements of a partition. Following the initial step of selecting TYPE, the algorithm considers other properties in $\mathcal{PO}$. Suppose $\langle \text{COLOUR} : blue \rangle$ is considered first. This property is true of e_2 and e_5. Since DF_2 refers to e_2, the new property can be added to M_{DF_2}. Since e_5 is not the sole referent of DF_1, the property induces a further partitioning of this fragment, resulting in a new DF. This is identical to DF_1 except that it refers only to e_5 and contains $\langle \text{COLOUR} : blue \rangle$. DF_1 itself now refers only to e_1. Once $\langle \text{COLOUR} : red \rangle$ is considered, it is added to the latter, yielding (4).

(4) DF_1 $\langle \{e_1\}, \langle \text{TYPE} : desk \rangle, \{\langle \text{COLOUR} : red \rangle\} \rangle$
    DF_2 $\langle \{e_2\}, \langle \text{TYPE} : sofa \rangle, \{\langle \text{COLOUR} : blue \rangle\} \rangle$
    DF_3 $\langle \{e_5\}, \langle \text{TYPE} : desk \rangle, \{\langle \text{COLOUR} : blue \rangle\} \rangle$

The procedure $updateDescription$, which creates and updates DFs, is formalised in Algorithm 2. When some property $\langle \text{A} : v \rangle$ is found to be 'useful' in relation to R (in a sense to be made precise), this function is called with two arguments: $\langle \text{A} : v \rangle$ itself, and $R' = [\![\langle \text{A} : v \rangle]\!] \cap R$, the referents of which $\langle \text{A} : v \rangle$ is true. The procedure iterates through the DFs in D_{part}, adding the property to any DF such that $R_{\text{DF}} \cap R' \neq \emptyset$, until R' is empty and all referents in it have been accounted for [2.2]. As indi-

cated in the informal discussion, there are two cases to consider for each DF:

1. $R_{\mathrm{DF}} \subseteq R'$ [2.4]. This corresponds to our example involving $\langle \mathrm{COLOUR} : blue \rangle$ and DF_2. The property is simply added to M_{DF} [2.5] and R' is updated by removing the elements thus accounted for [2.6].

2. $R_{\mathrm{DF}} \cap R' \neq \emptyset$ (but condition 1 does not hold) [2.7]. This occurred with $\langle \mathrm{COLOUR} : red \rangle$ in relation to DF_1. The procedure initialises R_{new}, a set holding those referents in R_{DF} which are also in R' [2.8]. A new DF (DF_{new}) is created, which is a copy of the old DF, except that (a) it contains the new property; and (b) its intended referents are R_{new} [2.9]. The new DF is included in the description [2.10], while the old DF is altered by removing R_{new} from R_{DF} [2.11]. This ensures that DFs denote disjoint subsets of R.

Two special cases arise when D_{part} is empty, or there are some elements of R' for which no DF exists. Both cases result in the construction of a new DF. An example of the former case is the initial state of the algorithm, when TYPE is added. As in example (3), the TYPE results in a new DF [2.16]. If a property is not a TYPE, the new DF has T set to null ($\perp$) and the property is included in M [2.18][5]. Note that this procedure easily generalises to the singular case, where D_{part} would only contain one DF.

4.2 Property selection criteria

IA$_{part}$'s content determination strategy maximises the similarity of a set by generating semantically parallel structures. Though contrastiveness plays a role in property selection, the 'preference' or conceptual salience of an attribute is also considered in the decision to propagate it across DFs.

Candidate properties for addition need only be true of at least one element of R. Because of the partitioning strategy, properties are not equally contrastive for all referents. Therefore, distractors are held in an associative array C, such that for all $r \in R$, $C[r]$ is the set of distractors for that referent at a given stage in the procedure. Contrastiveness is

---

[5]This only occurs if the KB is *incomplete*, that is, there some entities have no TYPE, so that R is not fully covered by the intended referents of the DFs when TYPE is initially added.

defined via the following Boolean function:

$$contrastive(\langle \mathrm{A} : v \rangle, R) \leftrightarrow$$
$$\exists r \in R : C[r] - [\![\langle \mathrm{A} : v \rangle]\!] \neq \emptyset \quad (5)$$

We turn next to salience and similarity. Let $A(D_{part})$ be the set of attributes included in D_{part}. A property is *salient* with respect to D_{part} if it satisfies the following:

$$salient(\langle \mathrm{A} : v \rangle, D_{part}) \leftrightarrow$$
$$\mathrm{A} \in A(D_{part}) \wedge (.713 \, p(\mathrm{A}, \mathrm{SG})^{.912} > 0.5) \quad (6)$$

that is, the attribute is already included in the description, and the predicted probability of its being propagated in more than one fragment of a description is greater than chance. A potential problem arises here. Consider the description in (3) once more. At this stage, IA$_{part}$ begins to consider COLOUR. The value *red* is true of e_1, but non-contrastive (all the desks which are not in R are red). If this is the first value of COLOUR considered, (6) returns `false` because the attribute has not been used in any part of the description. On later considering $\langle \mathrm{COLOUR} : blue \rangle$, the algorithm adds it to D_{part}, since it is contrastive for $\{e_2, e_5\}$, but will have failed to propagate COLOUR across fragments. As a result, IA$_{part}$ considers values of an attribute in order of *discriminatory power* (Dale, 1989), defined in the present context as follows:

$$\frac{|[\![\langle \mathrm{A} : v \rangle]\!] \cap R| + |[\![\langle \mathrm{A} : v \rangle]\!] - (U - R)|}{|[\![\langle \mathrm{A} : v \rangle]\!]|} \quad (7)$$

Discriminatory power depends on the number of referents a property includes in its extension, and the number of distractors $(U - R)$ it removes. By prioritising discriminatory values, the algorithm first considers and adds $\langle \mathrm{COLOUR} : blue \rangle$, and subsequently will include *red* because (6) returns `true`.

To continue with the example, at the stage represented by (4), only e_5 has been distinguished. ORIENTATION, the next attribute considered, is not contrastive for any referent. On considering SIZE, *small* is found to be contrastive for e_1 and e_2, and added to DF_1 and DF_2. However, SIZE is not added to DF_3, in spite of being present in two other fragments. This is because the probability function $p(\mathrm{SIZE}, \mathrm{PPS})$ returns a value below 0.5 (see Table 4, reflecting the relatively low conceptual salience of this attribute. The final description is *the blue desk, the small red*

		Mean	Mode	PRP
IA$_{bool}$	+ LOC	7.716	7	.7
	− LOC	8.335	7	3.5
IA$_{part}$	+ LOC	4.345	4	6.8
	− LOC	1.93	0	44.7

Table 5: Edit distance scores

desk and the small blue sofa. This example illustrates the limits set on semantic parallelism and similarity: only attributes which are salient enough are redundantly propagated across DFs.

5 Evaluation

IA$_{part}$ was compared to van Deemter's IA$_{bool}$ (§1) against human output in the evaluation sub-corpus PL$_2$ ($N = 405$). This was considered an adequate comparison, since IA$_{bool}$ shares with the current framework a genetic relationship with the IA. Other approaches, such as Gardent's (2002) brevity-oriented algorithm, would perform poorly on our data. As shown in §3, overspecification is extremely common in plural descriptions, suggesting that such a strategy is on the wrong track (but see §6).

IA$_{part}$ and IA$_{bool}$ were each run over the domain representation paired with each corpus description. The output logical form was compared to the LF compiled from the XML representation of an author's description (cf. Figure 1). LFs were represented as *and-or* trees, and compared using the tree edit distance algorithm of Shasha and Zhang (1990). On this measure, a value of 0 indicates identity.

Because only a subset of descriptions contain locative expressions, PL$_2$ was divided into a +LOC dataset ($N = 148$) and a −LOC dataset ($N = 257$). The preference orders for both algorithms were (C≫O≫S) for −LOC and (Y≫C≫X≫S≫O) for +LOC. These are suggested by the attribute probabilities in Table 4.

Table 5 displays the mean Edit score obtained by each algorithm on the two datasets, the modal (most frequent) value, and the *perfect recall percentage* (PRP), the proportion of Edit scores of 0, indicating perfect agreement with an author.

As the means and modes indicate, IA$_{part}$ outperformed IA$_{bool}$ on both datasets, with a consistently higher PRP (this coincides with the modal score in the case of −LOC). Pairwise $t-$tests showed that the trends were significant in both +LOC ($t(147) = 9.28$, $p < .001$) and −LOC ($t(256) = 10.039$, $p < .001$).

IA$_{bool}$ has a higher (worse) mean on −LOC, but

a better PRP than on +LOC. This apparent discrepancy is partly due to variance in the edit distance scores. For instance, because the Y attribute was highest in the preference order for +LOC, there were occasions when both referents could be identified using the same value of Y, which was therefore included by IA$_{bool}$ at first pass, before considering disjunctions. Since Y was highly preferred by authors (see Table 4), there was higher agreement on these cases, compared to those where the values of Y were different for the two referents. In the latter case, Y was only when disjunctions were considered, if at all. The worse performance of IA$_{part}$ on +LOC is due to a larger choice of attributes, also resulting in greater variance, and occasionally incurring higher Edit cost when the algorithm overspecified more than a human author. This is a potential shortcoming of the partitioning strategy outlined here, when it is applied to more complex domains.

(8) is an example of the algorithms' output, in a domain where COLOUR sufficed to distinguish the referents, which had different values of this attribute (i.e. an instance of the VDS condition). The formula returned by IA$_{part}$ (8a) is identical to the (LF of) the human-authored description (with Edit score of 0). The output of IA$_{bool}$ is shown in (8b).

(8)   (a) $(fan \land green) \lor (sofa \land blue)$

      (b) $(sofa \lor fan) \land small \land front \land (blue \lor green)$

As a result of IA$_{bool}$'s requiring a property or disjunction to be true of the the entire set of referents, COLOUR is not included until disjunctions are considered, while values of SIZE and ORIENTATION are included at first pass. By contrast, IA$_{part}$ includes COLOUR before any other attribute apart from TYPE. Though the data analysis suggests that overspecification is common in plural descriptions, IA$_{bool}$ overspecifies with the 'wrong' attributes (those which are relatively dispreferred compared to COLOUR). The rationale in IA$_{part}$ is to overspecify only if a property will enhance referent similarity, and is sufficiently salient. As for logical form, the Conjunctive Normal Form output of IA$_{bool}$ increases the Edit score, given the larger number of logical operators in (8b) compared to (8a).

6 Summary and conclusions

This paper presented an empirical study of plural reference, which showed that people undertake the dual strategy of partitioning sets based on the basic level TYPE of their elements and often redundantly

propagating attributes across disjuncts in adescription, modulo their salience. Our algorithm performs partitioning opportunistically, using KB information, and extends the notion of utility of a property beyond contrastiveness, utilising a statistical model of attribute salience. Evaluation results were positive, showing that these principles are on the right track. Though presented as a generalisation of Dale and Reiter's IA, the core aspects of the algorithm are more widely applicable. The partitioning strategy is related to a proposal by van Deemter and Krahmer (2006), which searches exhaustively for a partition of R whose elements can be described non-disjunctively. This differs from the present approach in that it is non-incremental and computationally costly.

The work described here highlights some open questions. In IA_{part}, a partition can be induced by any property. An alternative would be to aggregate same-type fragments of a description, producing NPs such as *the blue and red chairs* rather than *the blue chair and the red chair*. Limits on syntactic complexity of NPs are bound to play a role here, perhaps along lines suggested by Horacek (2004).

Though our data shows overspecification is often desirable, a preference order can make this excessive. In Table 1, a reference to all the entities *except* $\{e_1, e_2\}$ might use COLOUR, SIZE and ORIENTATION, rather than just *the large desks and sofas*. As we pointed out in §5, brevity-oriented alternatives, such as Gardent (2002), would perform poorly on our corpus data. Nevertheless, the balance between using salient attributes and being concise remains unclear. For example, the extent to which a property is shared among referents may be involved in the decision to use an otherwise dispreferred attribute (e.g. $\{e_3, \ldots, e_8\}$ are all of the same SIZE), suggesting that our notion of similarity could be extended, and taken beyond a preference-order based strategy.

Another way of simplifying descriptions involves negation (*the desks which are not red*) (Horacek, 2004). Though it can be handled relatively easily (van Deemter, 2002), there are several untackled empirical issues. In addition with those already raised, and the arguments made in this paper, they shown that many open questions remain despite more than a decade of intensive research in GRE. These questions are largely empirical.

References

A. Arts. 2004. *Overspecification in Instructive Texts*. Ph.D. thesis, University of Tilburg.

E. Belke and A. Meyer. 2002. Tracking the time course of multidimensional stimulus discrimination. *European Journal of Cognitive Psychology*, 14(2):237–266.

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

Robert Dale. 1989. Cooking up referring expressions. In *Proc. ACL-89*.

H. J. Eikmeyer and E. Ahlsèn. 1996. The cognitive process of referring to an object. In *Proc. 16th Scandinavian Conference on Linguistics*.

C. Gardent. 2002. Generating minimal definite descriptions. In *Proc. ACL-02*.

H. Horacek. 2004. On referring to sets of objects naturally. In *Proc. INLG0-04*.

G. L. Murphy. 2002. *The big book of concepts*. MIT Press, Cambridge, Ma.

T. Pechmann. 1989. Incremental speech production and referential overspecification. *Linguistics*, 27:89–110.

I. Rock. 1983. *The Logic of Perception*. MIT Press, Cambridge, Ma.

E. Rosch, C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439.

D. Shasha and K. Zhang. 1990. Fast algorithms for unit cost editing distance between trees. *Journal of Algorithms*, 11:581–621.

K. van Deemter and E. Krahmer. 2006. Graphs and booleans: On the generation of referring expressions. In H. Bunt and R. Muskens, editors, *Computing Meaning*, volume III. Kluwer Academic Publishers, Dordrecht.

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

I. van der Sluis, A. Gatt, and K. van Deemter. 2006. Manual for the TUNA corpus: Referring expressions in two domains. Technical report, University of Aberdeen.

M. Wertheimer. 1938. Laws of organization in perceptual forms. In W. Ellis, editor, *A Source Book of Gestalt Psychology*. Routledge & Kegan Paul.