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 descriptive 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 and incorporates corpus-derived heuristics. An exhaustive evaluation shows that the output closely matches human data.

1 Introduction

Generation of Referring Expressions (GRE) is a well-studied sub-task of microplanning in Natural Language Generation. Most algorithms in this area view GRE as a content determination problem, that is, their emphasis is on the construction of a semantic representation which is eventually mapped to a linguistic realisation (i.e. a noun phrase). Content Determination for 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. Under this view, reference is mainly about *identification* of an entitiy in a given context (represented by the KB), a well-studied pragmatic function of definite noun phrases in both the psycholinguistic and the computational literature (Olson, 1970).

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 co-

	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

ordinates. 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, is shown in Algorithm 1 (van Deemter, 2002). A non-disjunctive reference to a set R is possible just in case all the elements of R can be distinguished using the same attribute-value pairs. Such a description is equivalent to the logical conjunction of the properties in question. This algorithm, IA_{plur}, initialises a description D and a set of distractors C [1.1–1.2], and traverses an ordered list of properties, called the *preference order* ($\mathcal{PO}$) [1.3], which reflects general or domain-specific pref-

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

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

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 102–111, Prague, June 2007.

erences 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 has some *contrastive value*, that is, excludes some 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]. Given $R = \{e_1, e_2\}$, this algorithm would return the following description:

(1) ⟨ORIENTATION : *back*⟩ ∧ ⟨SIZE : *small*⟩

This description is overspecified, because ORIENTATION is not strictly necessary to distinguish the referents (⟨SIZE : *small*⟩ suffices). Moreover, the description does not include TYPE, though it has been argued that this is always required, as it maps to the head noun of an NP (Dale and Reiter, 1995). We will adopt this assumption here, for reasons explained below. Due to its hillclimbing nature, the IA avoids combinatorial search, unlike some predecessors which searched exhaustively for the briefest possible description of a referent (Dale, 1989), based on a strict interpretation of the Gricean Maxim of Quantity (Grice, 1975). Given that, under the view proposed by Olson (1970) among others, the function of a referential NP is to identify, a strict Gricean interpretation holds that it should contain no more information than necessary to achieve this goal.

The Incremental Algorithm constitutes a departure from this view given that it can overspecify through its use of a $\mathcal{PO}$. This has been justified on psycholinguistic grounds. Speakers overspecify their descriptions because they begin their formulation of a reference without exhaustively scanning a domain (Pechmann, 1989; Belke and Meyer, 2002). They prioritise the basic-level category (TYPE) of an object, and salient, absolute properties like COLOUR (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 that speakers conceptualise referents as *gestalts* (Pechmann, 1989) whose core is their basic-level TYPE (Murphy, 2002) and some other salient attributes like COLOUR. For instance, according to Schriefers and Pechmann (1988), an NP such as *the large black triangle* reflects a conceptualisation of the referent as a *black triangle*, of which the SIZE property is predicated. Thus, the TYPE+COLOUR combination is not mentally represented as two separable dimensions.

In what follows, we will sometimes refer to this principle as the *Conceptual Gestalts Principle*. 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 given the current state of the algorithm, whereas psycholinguistic research suggests that people include them irrespective of contrastiveness (but 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. Disjunction is required whenever elements of a set of referents R do not have identical distinguishing properties. For example, $\{e_1, e_3\}$ can be distinguished by the following Conjunctive Normal Form (CNF) description[1]:

(2) ⟨TYPE : *desk*⟩ ∧ (⟨COLOUR : *red*⟩ ∨ ⟨COLOUR : *blue*⟩) ∧ (⟨ORIENTATION : *right*⟩ ∨ ⟨ORIENTATION : *back*⟩)

Such a description would be returned by a generalised version of Algorithm 1 proposed by van Deemter (2002). This generalisation, IA_{bool} (so called because it handles all Boolean operators, such as negation and disjunction), first tries to find a non-disjunctive description using Algorithm 1. Failing this, it searches through disjunctions of properties of increasing length, conjoining them to the description. This procedure 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).

[1]Note that logical disjunction is usually rendered as linguistic coordination using *and*. Thus, *the table and the desk* is the union of things which are desks *or* tables.

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

Proposals to deal with (3) include Gardent's (2002) non-incremental, constraint-based algorithm to generate the briefest available description of a set, an approach extended in Gardent et al. (2004). 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 applicable to plurals.

This paper reports a corpus-based analysis of plural descriptions elicited in well-defined domains, of which Table 1 is an example. This study falls within a recent trend in which empirical issues in GRE have begun to be tackled (Gupta and Stent, 2005; Jordan and Walker, 2005; Viethen and Dale, 2006). We then propose an efficient algorithm for the generation of references to arbitrary sets, which combines corpus-derived heuristics and a partitioning-based procedure, comparing this to IA_{bool}. Unlike van Deemter (2002), we only focus on disjunction, leaving negation aside. Our starting point is the assumption that plurals, like singulars, evince preferences for certain attributes as predicted by the Conceptual Gestalts Principle. Based on previous work in Gestalt perception (Wertheimer, 1938; Rock, 1983), we propose an extension of this to sets, 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 determinant of logical 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:

(3) (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

These descriptions exemplify three possible sources of variation:

Disjunctive/Non-disjunctive: The last description, (3d), is non-disjunctive (i.e. it is logically a conjunction of properties). This, however, is only achievable through the use of a non-basic level value for the TYPE of the entities (*objects*). Using the basic-level would require the disjunction ($\langle$TYPE : *desk*$\rangle \vee \langle$TYPE : *sofa*$\rangle$), which is the case in (3a–c). Given that basic-level categories are preferred on independent grounds (Rosch et al., 1976), we would expect examples like (3d) to be relatively infrequent.

Aggregation: If a description is disjunctive, it may be aggregated, with properties common to all objects realised as wide-scope modifiers. For instance, in (3a), *small* modifies *desk and sofa*. By contrast, (3b) is *non-aggregated*: *small* occurs twice (modifying each coordinate in the NP). Non-aggregated, disjunctive descriptions are logically equivalent to a partition of a set. For instance, (3c) partitions the set $R = \{e_1, e_2\}$ into $\{\{e_1\}, \{e_2\}\}$, describing each element separately. Descriptions like (3b) are more overspecified than their aggregated counterparts due to the repetition of information.

Paralellism/Similarity: Non-aggregated, disjunctive descriptions (partitions) may exhibit *semantic parallelism*: In (3b), elements of the partition are described using exactly the same attributes (that is, TYPE, COLOUR, and SIZE). This is not the case in (3c), which does represent a partition but is *non-parallel*. Parallel structures maximise the similarity of elements of a partition, using the same attributes to describe both. The likelihood of propagation of an attribute across disjuncts is probably dependent on its degree of salience or preference (e.g. COLOUR is expected to be more likely to be found in a parallel structure than SIZE).

2 The data

The data for our study is a subset of the TUNA Corpus (Gatt et al., 2007), consisting 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 (see Table 1). 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

```
<DESCRIPTION num='pl'>
<DESCRIPTION num='sg'>
<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>
```

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

Figure 1: Corpus annotation examples

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. This markup abstracts away from differences in lexicalisation, making it an ideal resource to evaluate content determination algorithms, because it is *semantically transparent*, in the sense of this term used by van Deemter et al. (2006). This markup scheme also enables the compositional derivation of a logical form from a natural language description. For example, the XML representation of (3b) is shown in Figure 1, which also displays the LF derived from it. Each constituent NP in (3b) is annotated as a set of attributes enclosed by a DESCRIPTION tag, which is marked up as singular (sg). The two coordinates are further enclosed in a plural DESCRIPTION; correspondingly, the LF is a disjunction of (the LFs of) the two internal descriptions.

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) 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*).

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

	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

In terms of our introductory discussion, referents in *Value-Similar* conditions could be minimally distinguished using a conjunction of properties, while *Value-Dissimilar* referents required 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 $\langle \text{A} : v\rangle \vee \langle \text{A} : v'\rangle$. However, even in the VS condition, referents had different basic-level types. Thus, an author faced with a domain like Table 1 had at least the descriptive options in (3a–d). If they chose to refer to entities using basic-level values of TYPE, their description would be disjunctive (e.g. 3a). A non-disjunctive description would require the use of a superordinate value, as in (3d).

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. 3d) or disjunctive (3a–c). The latter were further classified into aggregated (3a) and non-aggregated (3b). 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 §1, 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 ob-

	Non-Parallel	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

jects. 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 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 authors are likely to redundantly propagate properties across disjuncts. This evidence goes against some recent proposals for plural reference generation which emphasise brevity (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 (3b) may be more likely than *non-parallel* ones (3c), 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

	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

descriptions into three classes, whereby a description is considered to be:

1. *underspecified* if it does not include a locative expression and omits some MD attributes;
2. *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;
3. *well-specified* otherwise.

Proportions of Parallel and Non-Parallel descriptions 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 (4), had a near-perfect fit to the data ($R^2 = .910$)[3]. 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 (4).

$$p(\text{A}, \text{PPS}) = .713\, p(\text{A}, \text{SG})^{.912} \tag{4}$$

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 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 if they are easily perceived and conceptualised. Thus, of the descriptions in (3) above, it is (3b) 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. First, it induces partitions opportunistically based on KB information, and this is is reflected in the way descriptions are represented. Second,, 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:

(5) DF$_1$ $\langle \{e_1, e_5\}, \langle \text{TYPE} : \textit{desk} \rangle, \emptyset \rangle$
DF$_2$ $\langle \{e_2\}, \langle \text{TYPE} : \textit{sofa} \rangle, \emptyset \rangle$

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 $\langle$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 $\langle$COLOUR : *blue*$\rangle$. DF$_1$ itself now refers only to e_1. Once $\langle$COLOUR : *red*$\rangle$ is considered, it is added to the latter, yielding (6).

(6) DF$_1$ $\langle \{e_1\}, \langle \text{TYPE} : \textit{desk} \rangle, \{\langle \text{COLOUR} : \textit{red} \rangle\} \rangle$
DF$_2$ $\langle \{e_2\}, \langle \text{TYPE} : \textit{sofa} \rangle, \{\langle \text{COLOUR} : \textit{blue} \rangle\} \rangle$
DF$_3$ $\langle \{e_5\}, \langle \text{TYPE} : \textit{desk} \rangle, \{\langle \text{COLOUR} : \textit{blue} \rangle\} \rangle$

The procedure $updateDescription$, which creates and updates DFs, is formalised in Algorithm 2. When some property $\langle$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: $\langle$A : $v\rangle$ itself, and $R' = [\![\langle \text{A} : v \rangle]\!] \cap R$, the referents of which $\langle$A : $v\rangle$ is true. The procedure iterates through

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

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

1: **for** $\langle R_{\text{DF}}, T_{\text{DF}}, M_{\text{DF}}\rangle \in D_{part}$ **do**
2:     **if** $R' = \emptyset$ **then**
3:         **return**
4:     **else if** $R_{\text{DF}} \subseteq R'$ **then**
5:         $M_{\text{DF}} \leftarrow M_{\text{DF}} \cup \{\langle \text{A} : v\rangle\}$
6:         $R' \leftarrow R' - R_{\text{DF}}$
7:     **else if** $R_{\text{DF}} \cap R' \neq \emptyset$ **then**
8:         $R_{new} \leftarrow R_{\text{DF}} \cap R'$
9:         $DF_{new} \leftarrow \langle R_{new}, T_{\text{DF}}, M_{\text{DF}} \cup \{\langle \text{A} : v\rangle\}\rangle$
10:         $D_{part} \leftarrow D_{part} \cup \{DF_{new}\}$
11:         $R_{\text{DF}} \leftarrow R_{\text{DF}} - R_{new}$
12:         $R' \leftarrow R' - R_{new}$
13:     **end if**
14: **end for**
15: **if** $\text{A} = \text{TYPE}$ **then**
16:     $D_{part} \leftarrow D_{part} \cup \{\langle R', \langle \text{A} : v\rangle, \emptyset\rangle\}$
17: **else**
18:     $D_{part} \leftarrow D_{part} \cup \{\langle R', \perp, \{\langle \text{A} : v\rangle\}\rangle\}$
19: **end if**

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

1. $R_{\text{DF}} \subseteq R'$ [2.4]. This corresponds to our example involving $\langle$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. Suppose $R_{\text{DF}} \not\subseteq R'$. If $R_{\text{DF}} \cap R'$ is empty, then $\langle \text{A} : v\rangle$ is not useful. Suppose on the other hand that $R_{\text{DF}} \cap R' \neq \emptyset$ [2.7]. This occurred with $\langle$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 (5), 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][4]. 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 constrastive for all referents. For instance, in (5), e_2 needs to be distinguished from the other sofas in Table 1, while $\{e_1, e_5\}$ need to be distinguished from the desks. 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 defined via the following Boolean function:

$$contrastive(\langle \text{A} : v\rangle, R) \leftrightarrow \exists r \in R : \; C[r] - [\![\langle \text{A} : v\rangle]\!] \neq \emptyset \quad (7)$$

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:

$$salient(\langle \text{A} : v\rangle, D_{part}) \leftrightarrow \text{A} \in A(D_{part}) \wedge (.713\, p(\text{A}, \text{SG})^{.912} > 0.5) \quad (8)$$

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 (5) once more. At this stage, IA$_{part}$ begins to consider COLOUR. The value *red* is true of e_1, but noncontrastive (all the desks which are not in R are red). If this is the first value of COLOUR considered, (8) returns `false` because the attribute has not been used in any part of the description. On later considering $\langle$COLOUR : *blue*$\rangle$, the algorithm adds it to

[4]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.

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:

$$\frac{|[\![\ \langle \text{A} : v\rangle\ ]\!] \cap R|\ + |[\![\ \langle \text{A} : v\rangle\ ]\!] - (U - R)|}{|[\![\ \langle \text{A} : v\rangle\ ]\!]|} \qquad (9)$$

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 $\langle$COLOUR : *blue*$\rangle$, and subsequently will include *red* because (8) returns `true`.

To continue with the example, at the stage represented by (6), 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 $p(\text{SIZE}, \text{PPS})$ 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 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.

4.3 Complexity

An estimate of the complexity of IA$_{part}$ must account for the way properties are selected (§4.2) and the way descriptions are updated (Algorithm 2).

Property selection involves checking properties for contrastive value and salience, and updating the ordering of values of each attribute based on discriminatory power (9). Clearly, the number of times this is carried out is bounded by the number of properties in the KB, which we denote n_p. Every time a property is selected, the discriminatory power of values changes (since the number of remaining distractors changes). Now, in the worst case, all n_p properties are selected by the algorithm [5]. Each time, the algorithm must compare the remaining properties pairwise for discriminatory power, a quadratic operation with complexity $O(n_p^2)$. With respect to the procedure $updateDescription$, we need to consider the number of iterations in the for loop starting at line [2.1]. This is bounded by $n_r = |R|$ (there can be no more DFs than there are referents). Once again, if at most n_p properties are selected, then the algorithm makes at most n_r iterations n_p times, yielding complexity $O(n_p n_r)$. Overall, then, IA$_{part}$ has a worst-case runtime complexity $O(n_p^3 n_r)$.

[5] Only unique properties need to be considered, as each property is selected at most once, though it can be included in more than one DF.

		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

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.

Some example outputs are given below, 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} (10a) is identical to the (LF of) the human-authored description (with Edit score of 0). The output of IA_{bool} is shown in (10b).

(10) (a) $(\textit{fan} \wedge \textit{green}) \vee (\textit{sofa} \wedge \textit{blue})$
'the green fan and the big sofa'
(b) $(\textit{sofa} \vee \textit{fan}) \wedge \textit{small} \wedge \textit{front} \wedge (\textit{blue} \vee \textit{green})$
'the small, blue and green sofa and fan'

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 overspecification is common in our data, IA_{bool} overspecifies with the 'wrong' attributes (those which are relatively dispreferred). 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 (10b) compared to (10a).

6 Summary and conclusions

This paper presented a study of plural reference, showing that people (a) partition sets based on the basic level TYPE or category of their elements and (b) redundantly propagate attributes across disjuncts in a description, modulo their salience. Our algorithm partitions a set opportunistically, and incorporates a corpus-derived heuristic to estimate the salience of a property. Evaluation results showed that these principles are on the right track, with significantly better performance over a previous model (van Deemter, 2002). The partitioning strategy is related to a proposal by van Deemter and Krahmer (2007), which performs exhaustive search for a partition of a set whose elements can be described non-disjunctively. Unlike the present approach, this algorithm is non-incremental and computationally costly.

IA_{part} initially performs partitioning based on the basic-level TYPE of objects, in line with the evidence. However, later partitions can be induced by other properties, possible yielding partitions even with same-TYPE referents (e.g. *the blue chair and the red chair*). Aggregation (*the blue and red chairs*) may be desirable in such cases, but limits on syntactic complexity of NPs are bound to play a role (Horacek, 2004). Another possible limitation of IA_{part} is that, despite strong evidence for overspecification, complex domains could yield very lengthy outputs. Strategies to avoid them include the utilisation of other boolean operators like negation (*the desks which are not red*) (Horacek, 2004). These issues are open to future empirical research.

7 Acknowledgements

Thanks to Ehud Reiter and Ielka van der Sluis for useful comments. This work forms part of the TUNA project (`www.csd.abdn.ac.uk/research/tuna/`), supported by EPSRC grant GR/S13330/01.

References

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

E. Belke and A. Meyer. 2002. Tracking the time course of multidimensional stimulus discrimination: Analysis of viewing patterns and processing times during same-different decisions. *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 *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, ACL-89.*

H. J. Eikmeyer and E. Ahlsèn. 1996. The cognitive process of referring to an object: A comparative study of german and swedish. In *Proceedings of the 16th Scandinavian Conference on Linguistics*.

C. Gardent, H. Manuélian, K. Striegnitz, and M. Amoia. 2004. Generating definite descriptions: Non-incrementality, inference, and data. In T. Pechman and C. Habel, editors, *Multidisciplinary Approaches to Language Production.* Mouton de Gruyter.

C. Gardent. 2002. Generating minimal definite descriptions. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL-02*.

A. Gatt, I. van der Sluis, and K. van Deemter. 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the 11th European Workshop on Natural Language Generation, ENLG-07.* To appear.

H.P. Grice. 1975. Logic and conversation. In P. Cole and J.L. Morgan, editors, *Syntax and Semantics: Speech Acts.*, volume III. Academic Press.

S. Gupta and A. J. Stent. 2005. Automatic evaluation of referring expression generation using corpora. In *Proceedings of the 1st Workshop on Using Corpora in NLG, Birmingham, UK.*

H. Horacek. 2004. On referring to sets of objects naturally. In *Proceedings of the 3rd International Conference on Natural Language Generation, INLG-04*.

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

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

D. R. Olson. 1970. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 77:257–273.

Thomas 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.

H. Schriefers and T. Pechmann. 1988. Incremental production of referential noun phrases by human speakers. In M. Zock and G. Sabah, editors, *Advances in Natural Language Generation*, volume 1. Pinter, London.

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. 2007. Graphs and booleans: On the generation of referring expressions. In H. Bunt and R. Muskens, editors, *Computing Meaning*, volume III. Springer, Berlin.

K. van Deemter, I. van der Sluis, and A. Gatt. 2006. Building a semantically transparent corpus for the generation of referring expressions. In *Proceedings of the 4th International Conference on Natural Language Generation (Special Session on Data Sharing and Evaluation), INLG-06.*

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

I. van der Sluis and E. Krahmer. 2005. Towards the generation of overspecified multimodal referring expressions. In *Proceedings of the Symposium on Dialogue Modelling and Generation, 15th Annual Meeting of the Society for Text and Discourse, STD-05.*

J. Viethen and R. Dale. 2006. Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of the 4th International Conference on Natural Language Generation, INLG-06.*

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