

Generation of Referring Expressions: Managing Structural Ambiguities *

Imtiaz Hussain Khan and Kees van Deemter and Graeme Ritchie

Department of Computing Science

University of Aberdeen

Aberdeen AB24 3UE, U.K.

{i.h.khan,k.vdeemter,g.ritchie}@abdn.ac.uk

Abstract

Existing algorithms for the Generation of Referring Expressions tend to generate distinguishing descriptions at the semantic level, disregarding the ways in which surface issues can affect their quality. This paper considers how these algorithms should deal with surface ambiguity, focussing on structural ambiguity. We propose that not all ambiguity is worth avoiding, and suggest some ways forward that attempt to avoid unwanted interpretations. We sketch the design of an algorithm motivated by our experimental findings.

1 Introduction

A Noun Phrase (NP) is a referring expression if its communicative purpose is to identify an object to a hearer. The Generation of Referring Expressions (GRE) is an integral part of most Natural Language Generation (NLG) systems (Reiter and Dale, 2000). The GRE task can informally be stated as follows. Given an intended referent (i.e., the object to be identified) and a set of distractors (i.e., other objects that can be confused with the referent), find a description that allows a hearer to identify its referent uniquely (Dale, 1992). Such a description is called a **Distinguishing Description** (DD). In practice, however, most GRE algorithms build sets of semantic properties available in a Knowledge Base (KB), rather than descriptions in natural language; surface issues are often ignored (exceptions are: (Stone and

Webber, 1998; Krahmer and Theune, 2002; Siddharthan and Copestake, 2004)). This is an important limitation, for example because *ambiguities* can be introduced in the step from properties to language descriptions. Such “surface ambiguities” take centerstage in this paper. More specifically, we shall be investigating situations where they lead to referential ambiguity, that is, unclarity as to what the intended referent of a referring expression is.

Example 1: Consider a scenario in which there are sheep and goats along with other animals, grazing in a meadow; some of the sheep and goats are black while others are either brown or yellow. Suppose our task is to single out the black sheep and black goats from the rest of the animals. Suppose an algorithm has generated the logical form¹ $(\text{Black} \cap \text{Sheep}) \cup (\text{Black} \cap \text{Goats})$, which could be realised as either *the black sheep and the black goats* or, more briefly, as *the black sheep and goats*. The latter NP expresses two non-equivalent logical formulae: (i) $(\text{Black} \cap \text{Sheep}) \cup \text{Goats}$, and (ii) $(\text{Black} \cap \text{Sheep}) \cup (\text{Black} \cap \text{Goats})$. Since both formulae correspond with a set of animals in the domain, referential ambiguity can result. On the other hand, *the black sheep and goats* is shorter and possibly more fluent. This example highlights the possible tension between brevity and lack of ambiguity. The question facing us in this paper is how to balance them.

This paper examines how GRE should deal with structural ambiguity, focussing on ambiguity of the form *the Adj Noun1 and Noun2*, also known as coordination ambiguity. We call referring expressions of this form **scopally ambiguous**, as the scope of *Adj* is unclear between wide scope (*Adj* applies to both nouns) and narrow scope (*Adj* applies only to *Noun1*).

¹In this paper, we use set-theoretic operators instead of logical connectives to represent logical forms.

* This work is supported by a University of Aberdeen Sixth Century Studentship, and EPSRC grant EP/E011764/1.

*© 2008. Licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported. Some rights reserved.

2 Approach

A cursory view of corpora such as the British National Corpus (BNC) reveals that there are many instances of coordination ambiguity:

1. *the black cats and dogs*
2. *the bearded men and women*
3. *the old men and women in the hats*

Psycholinguistic evidence suggests that, in many cases, these ambiguities could cause confusion for a hearer (Tanenhaus and Trueswell, 1995). Hence, it seems justifiable to have GRE avoid such kind of ambiguities. However, it also seems plausible that some readings may be very unlikely. For example, in (2) a wide-scope reading is, arguably, very unlikely. Abney and others have argued that every sentence is potentially ambiguous between many parses, even though we may not even notice this ambiguity (Abney, 1996; Wasow et al., 2005). This suggests that, in GRE as well, it might not be feasible to avoid all referential ambiguities all the time, and that the choice of referring expression should sometimes involve a balancing act in which *degree* of ambiguity is balanced against other properties of the generated expression, such as its length or fluency.

Building on earlier work by Inui et al. (Inui et al., 1992), Neumann (Neumann, 1994) suggested a general generate-parse-revise model for NLG, based on a reversible grammar. His generator generates a string which is then parsed to detect any structural ambiguities. If a string is found to be ambiguous then revision is used to produce an alternative, non-ambiguous string instead (if such a string exists). The likelihood of the different interpretations is not taken into account, however.

Our approach to the problem is to find out the likelihood of each interpretation of an NP, and to tailor GRE to avoid all **distractor interpretations** (i.e., interpretations that can be confused with the intended one) as suggested in (van Deemter, 2004). An interpretation can be confused with the intended one if it is more likely or almost as likely as the intended one. The problem is, how to determine the likelihood of different interpretations.

3 Getting likelihood from the BNC

In scopally ambiguous referring expressions, there is a tension between wide- and narrow-

scope interpretations. This can be viewed in terms of two competing forces: a *Coordination Force*, whereby *Noun1* and *Noun2* attract each other to form a syntactic unit, and a *Modification Force*, whereby *Adj* and *Noun1* attract each other to form a syntactic unit. Computational linguists have proposed using language corpora to estimate the likelihood of an interpretation (Wu and Furugori, 1998; Chantree et al., 2006). Chantree et al. used information from the Sketch Engine database (Kilgarriff, 2003) operating on the BNC to resolve coordination ambiguity. The Sketch Engine contains grammatical triples in the form of Word Sketches for each word, with each triple accompanied by a salience value indicating the likelihood of occurrence of the word with its argument in a grammatical relation. Word Sketches summarise the words' grammatical and collocational behavior.

Chantree et al. gathered a dataset of ambiguous phrases from a corpus of requirements specifications, and collected human judgements about their interpretations. They then used machine learning techniques combined with various heuristics to determine the most likely interpretation of a coordination. They identified two heuristics as particularly useful. One was the Coordination-Matches Heuristic: *if a coordination between two head nouns occurs (at all) within the corpus, then a wide-scope reading is likely*. The other was the Collocation-Frequency Heuristic: *if a modifier is collocated more frequently with the nearest head word than with the head word further away, then a narrow-scope reading is likely*. The best performance was achieved by combining the two heuristics: wide-scope reading is likely if Coordination-Matches heuristic gives a positive result and Collocation-Frequency heuristic gives a negative result. We decided to modify Chantree et al.'s approach in two ways and apply the modified approach to NLG.

Firstly, it seemed unlikely to us in the general case that the deciding factor is always whether two words co-occur *at all*. We therefore decided to separate cooccurrence percentages into ones that are very high and ones that are very low. Secondly, we observed that Chantree et al. take Coordination Force into account when they predict wide scope, but *not*

when they predict narrow scope. It would be more systematic – and more useful to an NLG system, which has to cope with all possible inputs – to consider all four combinations, of strong and weak, coordination and modification force. We define that there will be a Strong Coordination Force (SCF) if the collocational frequency between the two nouns is *high*, and a Weak Coordination Force (WCF) otherwise. Similarly, we define that there will be a Strong Modification Force (SMF) if the collocational frequency of *Adj* is *high* with *Noun1* and *low* with *Noun2*, and a Weak Modification Force (WMF) otherwise.

After a preliminary investigation of the data, we decided to operationalise high collocational frequency between two words as meaning that either of the two words appears among the top 30% collocates of the other word in a grammatical relation (of interest); low collocational frequency means that neither of the two words appears among the top 70% collocates of the other word in a grammatical relation. The hypotheses resulting from the above changes are investigated in the following section.

4 Empirical Studies

We conducted three experiments. The first two experiments ask what interpretation of a scopally ambiguous NP is the most plausible, thereby testing our generalisation of Chantree’s hypotheses. Knowing how an NP is interpreted is useful for an NLG system but not sufficient, because ambiguity needs to be traded off against other factors. For this reason, our third experiment asks which of several NPs are *preferred* by a reader.

4.1 Interpreting NPs

We use all four possible combinations of coordination and modification forces to predict an interpretation of a scopally ambiguous referring expression (see Table-1). An SMF would make a wide-scope reading highly unlikely (cf. (Wu and Furugori, 1998)). For instance, in *the bearded men and women* there is an SCF and an SMF, but in fact this phrase would be interpreted as a narrow-scope reading because of the scarcity of bearded women. On the other hand, a WMF could be in favor of a wide-scope reading. We expect that human readers would

opt for wide- and narrow-scope readings according to Table 1.

Table 1: Predicting an interpretation

Hypothesis 1: SCF \wedge SMF \Rightarrow NS
Hypothesis 2: SCF \wedge WMF \Rightarrow WS
Hypothesis 3: WCF \wedge SMF \Rightarrow NS
Hypothesis 4: WCF \wedge WMF \Rightarrow WS
WS: Wide scope; NS: Narrow scope

To test these hypotheses, we conducted two interpretation experiments, and rather than asking expert linguists to annotate the strings, we examined how ordinary readers interpret structurally ambiguous strings. In these experiments, given a referential domain and an English NP which attempts to identify a subset of objects in the domain, participants were asked to find the referent set of the NP.

4.1.1 Experiment 1

In this experiment, referential domains were constructed using real photographs of animals with some of the features printed alongside each photograph. Features were printed because 1) in a pilot study, we observed that some participants had difficulty in discerning some features in some of the photographs, and 2) we attribute some unusual features to some objects, e.g., we attributed cats with the feature BARKING although cats don’t bark in reality. Two pairs of nouns were used: one with SCF, and the other with WCF. For each pair of nouns, four different adjectives were used: two with SMF, and two with WMF. A trial in this experiment consists of a set of 9 pictures (placed in a 3 x 3 grid), and an English NP underneath these pictures. A sample trial is shown in Figure 1. Participants’ task was to remove the pictures (by mouse clicks on the pictures) that were referred to by the NP. A removed picture was immediately replaced by a blank rectangle (of the same size).

In each trial, we made sure that both wide- and narrow-scope readings are applicable. For example, for the instruction *Please, remove the red lions and horses*, in the domain there were 2 red lions, 2 red horses, and some (at least one) non-red horses. If a participant removes 2 red lions and 2 red horses, we count it as a wide-scope reading. However, if (s)he removes *all* the horses we count it as a narrow-scope reading. We also used 8 fillers, which do not

Figure 1: Interpreting an NP (using pictures)



Please, remove the roaring lions and horses.

contain a coordination in the NP (e.g., *the dogs on the left*). 60 self-reported native or fluent speakers of English, students from various UK universities, did the experiment on the web.²

Results and Discussion: Results were analysed according to whether a participant opted for a wide- or narrow-scope reading. The participants’ responses are shown in Table 2. A two-tailed sign binomial test was used to calculate statistical significance. The data indicate that word distribution information can reliably predict a wide-scope reading. However, our predictions for a narrow-scope reading are not confirmed. This may have been because of an intrinsic bias in favour of wide-scope interpretations. Another potential problem with the experiment is that some of the NPs shown to participants were rather unusual, involving bearded women, etc. Although the printed features underneath the pictures forced participants to take these unusual cases seriously, the clash between the picture (of a woman) and the printed feature (‘BEARDED’) that arose in such cases may have made participants’ responses unreliable. To avoid this problem we now turn to an experimental setup where we use Euler diagrams instead of iconic pictures.

4.1.2 Experiment 2

This experiment mirrors experiment 1, but we used Euler diagrams instead of pictures

²Here and in the other experiments reported in this paper, we ascertained that no important differences existed between the two groups of subjects. Focussing on Experiment 1, for example, no significant difference in the percentages of wide scope interpretations was found between native speakers and subjects who were merely fluent in English.

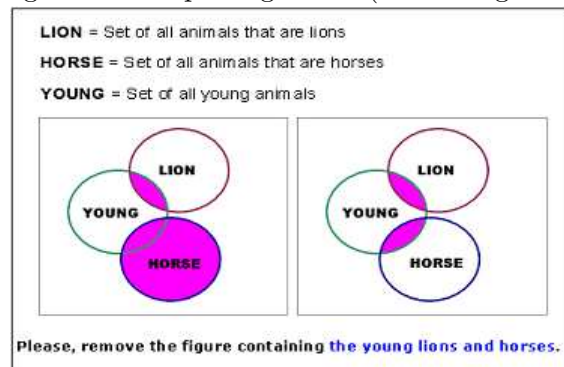
Table 2: Response proportions: Experiment 1

Force		PR	PJ	p-value
SCF	SMF	NS	NS (25/60)	0.52
SCF	WMF	WS	WS (57/60)	< 0.001
WCF	SMF	NS	NS (26/60)	0.12
WCF	WMF	WS	WS (53/60)	< 0.001

PR: Predicted Reading; PJ: Participants’ Judgement

to represent domain entities. Participants received a mini-tutorial on our version of Euler diagrams, where shaded areas denote the sets to which an NP might refer. The purpose of this tutorial was to make sure that the participants understand the semantics of these diagrams. A sample trial is shown in Figure 2 (where we expect that participants would remove the diagram on the right, which is counted as a wide-scope response). 60 self-reported native or fluent speakers of English, students from various UK universities, took part in this web-based experiment.

Figure 2: Interpreting an NP (Euler diagrams)



Results and Discussion: Results were recorded according to whether a participant opted for a wide- or narrow-scope reading. The participants’ responses are shown in Table 3. A two-tailed sign binomial test was used to calculate statistical significance of the results. This time, all four hypotheses are confirmed. We also observed, however, that in scopally ambiguous expressions, a narrow-scope reading tends to be particularly frequent in the *extreme* case where *Adj* has a zero co-occurrence with *Noun2* (in the BNC). We note that these results are in line with Chantree et al.

A critic might argue that the problem that was noted in connection with Experiment 1 applies to Experiment 2 as well, because it shows diagrams involving a “problematic” in-

Table 3: Response proportions: Experiment 2

Force		PR	PJ	p-value
SCF	SMF	NS	NS (51/60)	< 0.001
SCF	WMF	WS	WS (55/60)	< 0.001
WCF	SMF	NS	NS (46/60)	< 0.001
WCF	WMF	WS	WS (54/60)	< 0.001

tersection between, for example, BEARDED and WOMEN. The fact that women (arguably) cannot be bearded could cause subjects to reject these diagrams (choosing the other diagram instead, as in the diagram included in Fig. 3, which does not involve such an intersection). We would argue, however, that this does not cause an unwanted bias. The scarcity of bearded women is a legitimate reason for subjects to believe that a diagram that asserts their existence cannot be a proper interpretation of “bearded men and women”; it is just one of the many things that the corpus-based approach captures indirectly, without representing it explicitly. It is equally applicable to expressions like “handsome men and women”, where the corpus tells us that ‘handsome’ and ‘women’ do not go together well (even though one probably would not say they do not exist).

We have seen that Word Sketches can make reasonable predictions concerning the likelihood of the different interpretations of the NPs. But an NP that is *clear* (i.e., not likely to be misunderstood) may have other disadvantages. For example, it may lack fluency or it may be perceived as unnecessarily lengthy. For this reason, we also conducted an additional experiment in which we tested readers’ preferences.

4.2 Choosing the best NP

The question of how to choose between different NPs could be approached in a number of different ways: asking hearers which of several descriptions they prefer, asking hearers to rate several descriptions, measuring interpretation effort (time), measuring hearers’ errors etc.. We conducted a readers’ preference experiment where participants were asked to compare pairs of natural language descriptions of one and the same target set, selecting the one they found more appropriate. Brief descriptions took the form *the Adj Noun1 and Noun2*. Non-brief descriptions took the forms

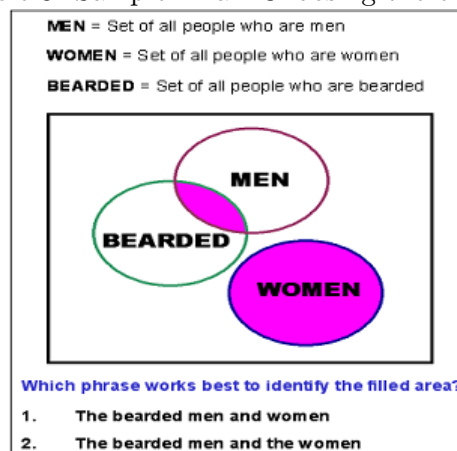
the Adj Noun1 and the Noun2 (for NS) and *the Adj Noun1 and the Adj Noun2* (for WS). A description is said to be clear if its predicted reading is the same as the intended one. By definition a non-brief description is always clear. Each description could either be brief or not ($\pm b$) and also clear or not ($\pm c$) (but not $(-b, -c)$, as this combination is not applicable in the present setting). We expected to find that:

Hypothesis 5: $(+c, +b)$ descriptions are preferred over ones that are $(+c, -b)$.

Hypothesis 6: $(+c, -b)$ descriptions are preferred over ones that are $(-c, +b)$.

4.2.1 Experiment 3

In this experiment, referential domains were represented using Euler diagrams. In each trial, participants were shown an Euler diagram, with some of its area filled to indicate the target referent. They were also shown two English NPs, which attempted to identify the filled area. A sample trial, where the intended reading is narrow scope, is shown in Figure 3. Each hypothesis was tested under two con-



ditions: 1) where the intended reading (IR) was WS; and 2) where the IR was NS. The 4 comparisons thus corresponded to 4 conditions (where PR stands for predicted reading):

C1. IR = WS & PR = WS

$(+c, +b)$ vs. $(+c, -b)$

C2. IR = NS & PR = NS

$(+c, +b)$ vs. $(+c, -b)$

C3. IR = WS & PR = NS

$(-c, +b)$ vs. $(+c, -b)$

C4. IR = NS & PR = WS

$(-c, +b)$ vs. $(+c, -b)$

46 self-reported native or fluent speakers of En-

glish, students from various UK universities, did the experiment on the web.

Results and Discussion: Results were coded according to whether a participant’s choice was $\pm b$ and/or $\pm c$. Table 4 displays response proportions. A two-tailed sign binomial test was used to calculate statistical significance of the results. The results confirm our hypotheses in all conditions, being highly statistically significant ($p < 0.001$).

Table 4: Response proportions: Experiment 3

	C1	C2	C3	C4
+b	91.3%	67.9%	26.1	14.5
+c	-	-	73.9%	88.5%

4.3 Summary of the Empirical Data

As hypothesised, Kilgarriff’s Word Sketches can be used to predict the most likely reading of a scopally ambiguous expression. It is also important to note that it is the Modification Force which is the deciding factor for a particular reading. Moreover, other things being equal, brief descriptions are preferred over longer ones. Since Experiment 2 (and, to an extent, Experiment 1) confirmed our hypotheses, we could have based our algorithm on these. As was noted in section 4.1.2, however, our data also suggest a slight modification of Hypotheses 1 and 3, because a preference for narrow scope existed mainly when the Adjective and the second Noun co-occurred very rarely. Therefore, we shall use a modified version of Strong Modification Force (SMF): SMF’ will mean that Adj and Noun2 have *zero* (rather than below 30%) cooccurrence in the BNC.

5 Applying results to GRE

In this section, we show how the results of the previous sections can be exploited in GRE. The patterns explored in the above correspond to *disjunctive* plural references. Disjunction is required whenever there is no conjunction of atomic properties that sets the elements of a set of referents apart from all the other objects in the domain. Recall example 1 (from §1), where the aim is to single out the black sheep and black goats from the rest of the animals. This task cannot be performed by a simple conjunction (i.e., of the form ‘the X’,

where X contains adjectives and nouns only), so disjunctions become unavoidable.

Various proposals have been made for allowing GRE algorithms to produce referring expressions of this kind (Stone, 2000; van Deemter, 2002; Gardent, 2002; Horacek, 2004). Here we take as our starting point the approach of (Gatt, 2007) (henceforth Gatt’s Algorithm with Partitioning or GAP). GAP is the only algorithm that produces a DD in Disjunctive Normal Form (DNF) while also guaranteeing that every “part” of the partition contains a noun. The DNF takes the form: $S_1 \cup S_2 \dots \cup S_n$, where each S_i itself expresses a conjunction of atomic properties. (For example, S_1 might be $\text{Sheep} \cap \text{Black}$, while S_2 is $\text{Goat} \cap \text{Black}$.) We sketch two extensions of this approach: the first, purely formal extension ensures that a *set* of such logical formulae is generated, rather than just one formula; all of these formulae are unambiguous, and logically equivalent with each other; but they all map to different strings of words. This is because we assume a very direct Linguistic Realisation strategy in which, for example, $((\text{Black} \cap \text{Sheep}) \cup \text{Goats})$ is worded as *the black sheep and goats*; syntactic ambiguity results from the lack of brackets in the English NP. The second, empirically based extension is to choose the “best” element of the set (of formulae) by making use of our experimental outcomes so as to balance clarity and brevity.

Since our predictions are based on words, we propose a model that constructs descriptions from words and in which the description building process is driven by words. We compute the *extension* (where the extension of a word w consists of all objects to which w applies) of a potentially ambiguous word by unifying the extensions of all its interpretations. Let p_1, p_2, \dots, p_n be the properties that a word w can express. Then the extension of w is:

$$[[w]] = \bigcup_{i=1}^{i=n} [[p_i]] \quad (1)$$

In what follows, a domain consists of a set \mathbb{D} of objects, and a set \mathbb{P} of properties applicable to objects in \mathbb{D} . Given a set of target referents $\mathbb{R} \subseteq \mathbb{D}$, the proposed algorithm will:

- lexicalise each $p \in \mathbb{P}$ into words; Lexicalisation takes a property as input and

returns the set of possible realisations of that property. For example, a property, say, AGED will be realised as (a set of) words $\{old, aged, senior\}$.

- build a DD in DNF using words, where the extension of a word is computed as indicated in equation 1. Each S_i must contain a head noun. For example, in the scenario presented in Example 1 under §1, it would produce a DD like: $(black \cap sheep) \cup (black \cap goats)$.
- apply transformation rules on the DD to construct a set of DDs that are logically equivalent to the DD. (See below.)
- realise each description in the set as English NPs using appropriate syntax. Each description is realised as one and only one NP, using the above realisation strategy.
- determine the most likely reading of each NP, by making use of Word Sketches.
- select the NP that is optimal given our empirical findings. (See below.)

Transformation Rules: In connection with reference to sets, it has been proposed to use the Q-M algorithm (McCluskey,) to find the *shortest* formula equivalent to a given input formula (van Deemter, 2002). In the present setting, the shortest formula might lead to a confusing NP after linguistic realisation. For example, the formula $Black \cap (Cats \cup Dogs)$ might be realised as *the black cats and dogs*, which could easily be misunderstood as $(Black \cap Cats) \cup Dogs$. For this purpose, we propose to use a set of transformation rules that allow us to find a set of formulae logically equivalent to the original formula; the aim is to make the set large enough that all the relevant expressive choices (as investigated in this paper) are represented. In particular, we need the following rules that operate on DNFS (where A is an adjective; B_1 and B_2 are nouns; X and Y are combinations of adjectives and nouns).

1. $((A \cap B_1) \cup (A \cap B_2)) \Rightarrow (A \cap (B_1 \cup B_2))$
2. $(X \cup Y) \Rightarrow (Y \cup X)$

After application of these transformation rules, the original description φ (i.e., the formula produced by an algorithm such as GAP) is replaced by a *set* of formulae \mathbb{F} all of whose elements are logically equivalent to φ . The elements of \mathbb{F} are then realised as NPs. The *clarity* of each NP is determined as follows (where

PR and IR stand for predicted reading and intended reading, respectively).

```

If SMF' then PR is NS
Else If WMF then PR is WS
Else PR is {NS, WS}
EndIf
If (PR = IR) then NP is clear
Else NP is unclear
EndIf

```

If, after transformations, several of the resulting descriptions are clear then the choice between them needs to be taken on other grounds. To do this, we give preference to the shortest of all descriptions that are clear (measured in terms of number of words in the NP). If ties still arise then we suggest that fluency is taken into account, for example by preferring NP whose structure is most frequent in the BNC. This procedure will often result in NPs that are ‘clear’ even though they are syntactically ambiguous.

Example 2: Let the domain be represented as: $\{MAN(e_1, e_2, e_6), WOMAN(e_3, e_4, e_5), YOUNG(e_5, e_6), OLD(e_1, e_2, e_3, e_4)\}$. Our task is to single out $\{e_1, e_2, e_3, e_4\}$ from rest of the entities. First, properties are lexicalised into words. Suppose the relevant words are the ones in the list $\mathbb{Q} = \langle man, woman, old, young \rangle$. Then, the algorithm takes each word $w \in \mathbb{Q}$ in turn and constructs a DD: $(old \cap man) \cup (old \cap woman)$. The transformation rules then produce $\{old \cap (man \cup woman), old \cap (woman \cup man), (old \cap man) \cup (old \cap woman), (old \cap woman) \cup (old \cap man)\}$. These formulae are realised as: (1) *the old men and women*, (2) *the old women and men*, (3) *the old men and the old women* and (4) *the old women and the old men*. The NPs (1) and (2) are structurally ambiguous, but the Word Sketches rule out the unintended reading of both NPs (with narrow scope for the adjective), so they are both clear. The NPs (3) and (4) are structurally unambiguous. All NPs are therefore clear, but (1) and (2) are preferred because they are shorter than (3) and (4). Corpus frequency suggests that the tie between (1) and (2) is resolved by opting for the more frequent pattern (1).

6 Conclusions and future work

We highlighted that structural ambiguity, which is often ignored in the GRE could cause

confusion for a hearer and, therefore, should be dealt with. Based on psycholinguistic evidence that avoidance of all ambiguity is hard, we suggested an approach that avoids referring expressions that have distractor interpretations. We did: (1) interpretation experiments and found that Word Sketches can be used to make distractor interpretation precise; and (2) an experiment with human readers that trades-off clarity and brevity. A GRE algorithm is sketched that balances these factors based on our experimental findings.

We aim to extend this work in two directions. First, we hypothesise that our approach can help NLG systems handle other surface ambiguities, for instance involving PP-attachment. Second, we realise that contextual factors are likely to affect people’s interpretive and generative inclinations. Therefore, in light of the work reported in this paper, it would be interesting to explore the effect of co-occurrences in a given text upon the interpretation of NPs occurring later *in that same text*, since the effect of such earlier occurrences on readers’ interpretation could conceivably ‘drown out’ the generic likelihoods based on Word Sketches that have formed the main subject matter of this paper.

References

- Abney, S. 1996. Statistical methods and linguistics. In Klavans, Judith and Philip Resnik, editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pages 1–26. The MIT Press, Cambridge, Massachusetts.
- Chantree, F., B. Nuseibeh, A. de Roeck, and A. Willis. 2006. Identifying nocuous ambiguities in requirements specifications. In *Proceedings of 14th IEEE International Requirements Engineering conference*, Minnesota, U.S.A.
- Dale, R. 1992. *Generating Referring Expressions: Building Descriptions in a Domain of Objects and Processes*. MIT Press.
- Gardent, C. 2002. Generating minimal definite descriptions. In *Proceedings of the 40th Annual Meeting of the ACL*, Philadelphia, USA.
- Gatt, A. 2007. *Generating Coherent References to Multiple Entities*. Ph.D. thesis, University of Aberdeen, Aberdeen, Scotland.
- Horacek, H. 2004. On referring to sets of objects naturally. In *Proceedings of the 3rd International Conference on NLG*, pages 70–79, UK.
- Inui, K., T. Tokunaga, and H. Tanaka. 1992. Text revision: A model and its implementation. In *Proceedings of the 6th International Workshop on NLG*, pages 215–230, Berlin, Heidelberg.
- Kilgarriff, A. 2003. Thesauruses for natural language processing. In *Proceedings of NLP-KE*, pages 5–13, Beijing, China.
- Krahmer, E. and M. Theune. 2002. Efficient context-sensitive generation of referring expressions. In van Deemter, K. and R. Kibble, editors, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, *CSLI Publications*, pages 223–264.
- McCluskey, E. J. *Introduction to the Theory of Switching Circuits*. McGraw-Hill Book Co.
- Neumann, G. 1994. *A Uniform Computational Model for Natural Language Parsing and Generation*. Ph.D. thesis, University of the Saarland.
- Reiter, E. and R. Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.
- Siddharthan, A. and A. Copestake. 2004. Generating referring expressions in open domains. In *Proceedings of the 42nd Annual Meeting of the ACL*, Barcelona, Spain.
- Stone, M. and B. Webber. 1998. Textual economy through close coupling of syntax and semantics. In *Proceedings of the 9th International Workshop on NLG*, pages 178–187, New Brunswick, New Jersey.
- Stone, M. 2000. On identifying sets. In *Proceedings of the 1st INLG Conference*, pages 116–123, Mitzpe Ramon.
- Tanenhaus, M.K. and J.C. Trueswell. 1995. Sentence comprehension. In Miller, J. and P. Eimas, editors, *Handbook of Perception and Cognition, Vol. 11: Speech, Language and Communication*, pages 217–262. New York: Academic Press.
- van Deemter, K. 2002. Generating referring expressions: Boolean extensions of the incremental algorithm. *Comp. Linguistics*, 28(1):37–52.
- van Deemter, K. 2004. Towards a probabilistic version of bidirectional OT syntax and semantics. *Journal of Semantics*, 21(3):251–281.
- Wasow, T., A. Perfors, and D. Beaver. 2005. The puzzle of ambiguity. In Orgun, O. and P. Sells, editors, *Morphology and The Web of Grammar: Essays in Memory of Steven G. Lapointe*. CSLI Publications.
- Wu, H. and T. Furugori. 1998. A computational method for resolving ambiguities in coordinate structures. In *Proceedings of PACLIC-12*, pages 263–270, National University of Singapore.