Spatial descriptions as referring expressions in the MapTask domain

Sebastian Varges
Information Technology Research Institute
University of Brighton
Sebastian.Varges@itri.brighton.ac.uk

Abstract

We discuss work-in-progress on a hybrid approach to the generation of spatial descriptions, using the maps of the Map Task dialogue corpus as domain models. We treat spatial descriptions as referring expressions that distinguish particular points on the maps from all other points (potential 'distractors'). Our approach is based on rule-based overgeneration of spatial descriptions combined with ranking which currently is based on explicit goodness criteria but will ultimately be corpus-based. Ranking for content determination tasks such as referring expression generation raises a number of deep and vexing questions about the role of corpora in NLG, the kind of knowledge they can provide and how it is used.

1 Introduction, or: The lack of domain model annotation in corpora used for ranking in NLG

In recent years, ranking approaches to Natural Language Generation (NLG) have become increasingly popular. They abandon the idea of generation as a deterministic decision-making process in favour of approaches that combine overgeneration with ranking at some stage in processing. A major motivation is the potential reduction of manual development costs and increased adaptability and robustness.

Several approaches to sentence realization use ranking models trained on corpora of human-authored texts to judge the fluency of the candidates produced by the generation system. The work of [Langkilde and Knight, 1998; Langkilde, 2002] describes a sentence realizer that uses word ngram models trained on a corpus of 250 million words to rank candidates. [Varges and Mellish, 2001] present an approach to sentence realization that employs an instance-based ranker trained on a semantically annotated subset of the Penn treebank II ('Who's News' texts). [Ratnaparkhi, 2000] describes a sentence realizer that had been trained on a domain-specific corpus (in the air travel domain) augmented with semantic attribute-value pairs. [Bangalore and Rambow, 2000] describe a realizer that uses a word ngram model combined with a tree-based stochastic model trained on a version of the Penn treebank annotated in XTAG grammar format. [Karamanis *et al.*, 2004] discuss centering-based metrics of coherence that could be used for choosing among competing text structures. The metrics are derived from the Gnome corpus [Poesio *et al.*, 2004].

In sum, these approaches use corpora with various types of annotation: syntactic trees, semantic roles, text structure, or no annotation at all (for word-based ngram models). However, what they all have in common, even when dealing with higher-level text structures, is the absence of any domain model annotation, i.e. information about the available knowledge pool from which the content was chosen. This seems to be unproblematic for surface realization where the semantic input has been determined beforehand.

This paper asks what the lack of domain information means for ranking in the context of content determination, focusing on the generation of referring expressions (GRE). A particularly intriguing aspect of GRE is the role of *distractors* in choosing the content (types, attributes, relations) used to describe the *target object(s)*. For example, we may describe a target as 'the red car' if there is also a blue one, but we may just describe it as 'the car' if there are no other cars in the domain (but possibly objects of other types). [Stone, 2003] proposes to use this observation to reason backwards from a given referring expression to the state of the knowledge base that motivated it. We may call this the 'presuppositional' or 'abductive' view of GRE. The approach is intended to address the knowledge acquisition bottleneck in NLG by means of example specifications constructed for the purpose of knowledge acquisition. It seems to us that, if the approach were to be applied to actual text corpora, one needed to address the fact that people often include 'redundant' attributes that do not eliminate any distractors. Thus, 'the red car' does not necessarily presuppose the existence of another car of different colour. Furthermore, there are likely to be a large number of domain models/knowledge bases that could have motivated the production of a referring expression.

[Siddharthan and Copestake, 2004] take a corpus-based perspective and essentially regard a text as a knowledge base from which descriptions of domain objects can be extracted. Some NPs are descriptions of the same object (for example if they have the same head noun and share attributes and relations in certain ways), others are deemed distractors. It seems that, in contrast to [Stone, 2003], this approach cannot recover those domain objects or properties that are never mentioned because it only extracts what is explicitly stated in the text.

Both the work reported in [Stone, 2003] and in [Siddharthan and Copestake, 2004] can be seen as attempts to deal with the lack of domain model information in situations where only the surface forms of referring expressions are given. Obtaining such a domain model is highly desirable in order to establish which part of a larger knowledge pool is actually selected for realization. This could be used to automatically learn models of content selection, for example. However, as we observed above, most corpora do not provide this kind of knowledge for obvious practical reasons. For example, how can we know what knowledge a Wall Street Journal author had available at the time of writing?

In this paper, we describe work-in-progress on exploiting a corpus that provides not only surface forms but also domain model information: the MapTask dialogue corpus [Anderson *et al.*, 1991].

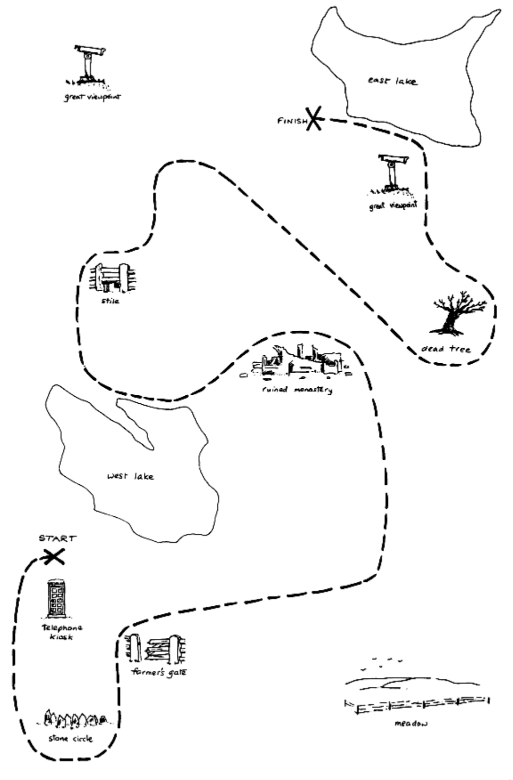

Figure 1: A 'giver' map of the MapTask experiments

2 Spatial descriptions as referring expressions

In the Map Task dialogues, a subject gives route directions to another subject, involving the production of descriptions such as 'at the left-hand side of the banana tree' and 'about three quarters up on the page, to the extreme left'. 16 Maps and 32 subjects (8 groups of 4 speakers) were used to produce 128 dialogues. The subjects were not able to see each other's maps and thus had to resort to verbal descriptions of the map contents. There are some (intentional) mismatches the subject's maps such as additional landmarks, changed names etc. This is a good source of spatial descriptions, for example: 'have you got gorillas? ... well, they're in the bottom left-hand corner.'

Figure 1 shows one of the 'giver' maps, which, in contrast to the corresponding 'follower' map, shows the route the follower is supposed to take. A main characteristic of both giver and follower maps is the display of *named* landmarks. These names typically refer to the type of the landmark. With a few exceptions, for example the 'great viewpoint' in figure 1, most of the landmark names only occur once. This seems to make it difficult to use the MapTask dialogues from the perspective of GRE: the names/types rule out most distractors and there is not much left to do for a GRE algorithm. However, as can be seen in figure 1, the routes do not lead through the landmarks but rather around them along feature-less points. The subjects of the MapTask experiments therefore often refer to points on the route, for example to those where the next turn had to be taken.

These observations can be used to frame the generation of spatial descriptions as a GRE task in which target points on the map are distinguished from all other points (the distractors). Since most points are feature-less, two of the properties commonly used in GRE, types and attributes, cannot be used in many cases. This leaves the third property, relations, which can be used by a GRE algorithm to relate the target position to surrounding landmarks on the maps. This is also what the subjects in the MapTask experiments are doing.

Our current, on-going work addresses the generation of descriptions referring to individual points on the maps. Ultimately, we hope to move on to the generation of descriptions of (straight) paths encompassing start and end points, and a description of how to travel between these. Looking at the MapTask corpus from the perspective of GRE, we make the following observations:

- The MapTask corpus consists of transcriptions of spoken language and contains many disfluencies. A spatial description can even span more than one turn, for example: 'okay ... fine move ... upwards ... an– ... so that you're to the left of the broken gate .' TURN 'and just level to the gatepost .' We expect more polished, written language as output of our generator.
- GRE only deals with a subset of the corpus. We need to find ways of making use of the appropriate parts of the data while ignoring the other ones.

- Spatial descriptions containing some form of vagueness seem to be frequent:[1] ‘quite close to the mountain on its right-hand side’, ‘just before the middle of the page’, ‘a bit up from the fallen pillars on the left’, ‘about two inches southwest’. There even seem to be rare cases of vague types: ‘a sort of curve’.
- Discourse context is of crucial importance, i.e. many spatial descriptions do not mention a particular point for the first time. Furthermore, already established information is not always given explicitly, for example ‘two inches to the left [from where I am now].’
- Perspective is a related issue: it seems that the subjects zoom in on parts of the map, ignoring for example a second lake at the other end of the map.
- The subjects switch between referring to objects from an inside-the-map perspective (‘beneath the tree’) to using the physical maps as a perspective (‘a couple of centimeters ... from the bottom of the page’). The latter may be caused by the lack of a grid indicating the distances in miles or kilometers.

In sum, the MapTask corpus contains a wealth of data combined with a domain model (the maps). The challenge is to make best use of these resources. In the following sections, we report on work-in-progress on hybrid rule-based/instance-based generation of spatial referring expressions.

3 Overgenerating spatial descriptions

Following the inferential approach to GRE [Varges, 2004; Varges and van Deemter, 2005], we are implementing a system that finds all combinations of properties that are true of a given target referent and distinguish it from its distractors. The approach pairs the logical forms derived from a domain representation with the corresponding ‘extensions’, the set of objects that the logical form denotes. We represent spatial domains as grids and domain objects (landmarks) as sets of coordinates of those grids. For example, the telephone kiosk in figure 1 is represented as the set $\{(2,18),(2,19)\}$. The grid resolution has implications for the definition of targets, which are, in fact, target areas, i.e. they often consist of more than one coordinate.

We implemented a number of content determination rules that recursively build up descriptions, starting from (NPs realizing) the landmarks of the domain model:

1. spatial functions: `Prep NP` → `PP`: ‘above the west lake’. The spatial functions used so far are: ‘above’, ‘below’, ‘to the left of’, ‘to the right of’.
2. intersection: `PP and PP` → `PP`: ‘above the west lake and to the left of the great viewpoint’.
3. union: `NP or NP` → `NP`: ‘the stile or the ruined monastery’.

[1]The following examples do not always refer to the map shown in figure 1.

```
+-----------------+
|                 |   grid resolution: 17 x 24
|          x    x |   (last two rows omitted from grid)
|   x      xxxxxx |
|   x       xxxxx |
|           xxxxx |
|            x  x |
|            x    |
|                 |
|    xx           |   This map is the ‘extension’ of
|    xx        x  |   ‘the telephone kiosk or the great
|              x  |    viewpoint or the ruined monastry
|        xxxx     |    or the stone circle or the dead
|  xxxxxx         |    tree ...’
|   xxxxx         |
|   xxxxx         |
|    xxxx         |
|       x         |
|                 |
| #x              |   #: target area of example
| #x  xx          |       (see section 4)
|     xx    xxxxx |
|  xxx      xxxxx |
|                 |
+-----------------+
```

Figure 2: Grid representation of map in fig. 1

The descriptions generated by these rules are all associated with an extension. For example ‘above X’ denotes all all the points above X (this may be changed to those points that are also ‘close’ to X). The grid in figure 2 is a graphical depiction of an extension associated with the disjoined NPs shown on the right. All generated descriptions can be visualized in this way.

The three content determination rules listed above are not sufficient for singling out all areas of the maps. For example, the corners of the maps are typically not ‘reachable’. Therefore, we define a further rule:

4. recursive spatial functions: `Prep PP` → `PP`: ‘above the points to the left of the farmer’s gate’.

We intend to generate a wide variety of realization candidates. For example, rule 4 could also produce the more fluent (and realistic) ‘above the farmer’s gate slightly to the left’, which will also require us to make vagueness more precise. An alternative is to use non-recursive PPs like ‘southeast of X’.

4 Ranking spatial descriptions

As candidates for ranking we use those NPs and PPs that contain at least one target coordinate, a loose definition that increases the number of candidates. We always generate NP ‘the points’ as a candidate which refers to all coordinates. However, it should be ranked low because it does not rule out any distractors. We currently use the ratio of extension size to described target coordinates as our first (non-empirical) ranking criterion (‘e/t’ below). The second (non-empirical) criterion is brevity, measured by the number of characters (‘chars’). Here are some ranked output candidates for the target area labeled '#' in figure 2 (which contains two points):

```
e/t | chars | realization          |  extension size
-------------------------------------------------------
2   34  to the left of the telephone kiosk       4
2   71  to the left of the farmer's gate
         and to the left of the telephone kiosk  2
2   88  above the points to the left of the
         farmer's gate and to the left of the
         telephone kiosk                         2
10  32  to the left of the farmer's gate        10
225 10  the points                             450
```

The first candidate is preferred because it has the same e/t ratio as some of its competitors but in addition is also shorter than these. In fact, this is how one of the subjects refers to the starting point in one of the dialogues. The third candidate requires the use of appropriate bracketing to yield the desired reading. For example, the generator could introduce a comma after 'gate'.

5 Toward using empirical data for ranking

The generation rules sketched above produce non-redundant spatial descriptions, i.e. the generator is 'economical' [Stone, 2003] and follows the 'local brevity' interpretation of the Gricean Maxims [Dale and Reiter, 1995]. The candidate of least 'complexity' is the 'full brevity' solution. A word similarity-based ranker could align the generation output (i.e. the highest-ranked candidate) with previous utterances in the discourse context. To increase choice, we intend to also generate additional candidates that include a limited amount of redundant information. One could furthermore generate candidates that, by themselves, do not rule out all distractors. In contrast to the inclusion of redundant information, these candidates would only be safe to use in combination with, for example, a reliable model of discourse salience that reduces the set of possible distractors.

It is possible (but not without difficulty) to annotate parts of the corpus with map coordinates. For example, we can annotate the turn 'on the right side of the tree' with coordinates (15,9), (15,10) in figure 2. Further markup could be applied to 'redundant' information (in the GRE sense) or highlight available discourse context. However, for obvious reasons it is preferable to use corpus data without any additional annotation for ranking. The maps enable us to determine how much we gain from the availability of a domain model.

6 Acknowledgments

Our thanks go to Kees van Deemter, Richard Power and Albert Gatt for very helpful discussions on the topics of this paper. The presented research has been conducted as part of the TUNA[2] project funded by EPSRC (grant number GR/S13330/01).

[2] Towards a UNified Algorithm for the Generation of Referring Expressions.

References

[Anderson *et al.*, 1991] A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert. The HCRC Map Task Corpus. *Language and Speech*, 34:351–366, 1991.

[Bangalore and Rambow, 2000] Srinivas Bangalore and Owen Rambow. Corpus-based Lexical Choice in Natural Language Generation. In *Proc. of ACL-00*, 2000.

[Dale and Reiter, 1995] Robert Dale and Ehud Reiter. Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19:233–263, 1995.

[Karamanis *et al.*, 2004] Nikiforos Karamanis, Massimo Poesio, Chris Mellish, and Jon Oberlander. Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *Proc. of ACL-04*, 2004.

[Langkilde and Knight, 1998] Irene Langkilde and Kevin Knight. Generation that Exploits Corpus-based Statistical Knowledge. In *Proc. of COLING/ACL-98*, 1998.

[Langkilde, 2002] Irene Langkilde. An Empirical Verification of Coverage and Correctness for a General-Purpose Sentence Generator. In *Proc. of INLG-02*, 2002.

[Poesio *et al.*, 2004] Massimo Poesio, Rosemary Stevenson, Barbara di Eugenio, and Janet Hitzeman. Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3), 2004.

[Ratnaparkhi, 2000] Adwait Ratnaparkhi. Trainable Methods for Surface Natural Language Generation. In *Proc. of NAACL-00*, 2000.

[Siddharthan and Copestake, 2004] Advaith Siddharthan and Ann Copestake. Generating referring expressions in open domains. In *Proc. of ACL-04*, 2004.

[Stone, 2003] Matthew Stone. Specifying Generation of Referring Expressions by Example. In *Proc. of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, 2003.

[Varges and Mellish, 2001] Sebastian Varges and Chris Mellish. Instance-based Natural Language Generation. In *Proc. NAACL-01*, 2001.

[Varges and van Deemter, 2005] Sebastian Varges and Kees van Deemter. Generating referring expressions containing quantifiers. In *Proceedings of IWCS-6*, 2005.

[Varges, 2004] Sebastian Varges. Overgenerating referring expressions involving relations. In *Proc. of INLG-04*, 2004.