Quantifying Spatial Prepositions: an Experimental Study

Mark M. Hall
School of Computer Science
Cardiff University
5 The Parade, Roath
Cardiff CF24 3AA, United Kingdom
M.M.Hall@cs.cardiff.ac.uk

Christopher B. Jones
School of Computer Science
Cardiff University
5 The Parade, Roath
Cardiff CF24 3AA, United Kingdom
C.B.Jones@cs.cardiff.ac.uk

ABSTRACT

Many aspects of spatial language concerned with relationships between spatial entities are essentially vague. Current GIS technology provides very little support for dealing with this vagueness, partially because there is a lack of quantitative data and models for vague spatial relations. This paper presents an experiment that looks at quantifying spatial prepositions. In the context of image captions, the cardinal directions are analysed in an existing set of image captions, with respect to the spatial distribution of the locations of the target object (figure) and the reference object (ground). Future work will focus on using these results to improve current GIS solutions in a wide variety of scenarios.

Categories and Subject Descriptors

J.2 [**Physical Sciences and Engineering**]: [Earth and atmospheric sciences]; H.1.2 [**Models and Principles**]: User / Machine Systems—*Human information processing*

General Terms

Spatial language, experimental study

Keywords

Spatial language, experimental study, vagueness, quantitative modelling, cognitive geography

1. INTRODUCTION

Many aspects of spatial language concerned with real world features and with the relationships between them are essentially vague. While this vagueness is managed quite effectively in natural language communication between people, there are currently only very limited facilities for interpreting such language when used to communicate with computers. Most geographical information systems (GIS) provide support for interpretation of relatively precise and unambiguous spatial terminology within structured queries relating to distance-based search and to Boolean topological spatial operators like inside, crosses and meets. This is adequate for many professional applications of GIS technology, but there are many applications of public information systems, as well as some professional GIS, where it would be desirable to be able to make "intelligent" interpretations of vague spatial language. This is the case for example when searching for information relating to imprecisely defined places such as the "Mid West" or the "South of France", and when using vague spatial prepositions such as "north-of". In addition to processing requests for information it would be desirable to be able to interpret the use of such language when employed in textual descriptions of the location of phenomena such as archaeological, botanical or geological samples and the descriptions of events provided by emergency services.

In the present study we analyse the use of spatial prepositions for the purposes of captioning photographic images. Typically spatial propositions relate at least two objects to each other as in "A pond north of Stackpole". In this paper the object that acts as the reference object, "Stackpole" in this case, will be referred to as the *ground*, while the referred object "pond" will be called the *figure*. In the image captions employed in our study, the *figure* object usually describes the content of the image, while the *ground* is referred to by the name (toponym) of a place in the close vicinity. Using knowledge of where photographs were taken and the georeferences of the ground place names in a caption, we mapped the spatial distribution of the figure locations relative to the ground, for multiple uses of each of several frequently employed spatial prepositions. Clearly the interpretation of a spatial proposition may vary according to the context, in particular the scale of the ground object. Here the ground objects were "inhabited places" that fell mostly within a confined range of size and hence provided a reasonably consistent context for the use of the preposition.

2. SPATIAL RELATIONS AND VAGUENESS

2.1 Spatial Language

Spatial language consists primarily of objects located in space and spatial prepositions that describe spatial relations between these objects [7]. In the kind of image caption language this paper focuses on, the role of the located object is often taken by places represented by their toponyms. Compared to the number of located objects, toponyms and the number of spatial configurations that are possible between these, the number of spatial prepositions in frequent use is very small. Thus to be able to describe all possible configu-

ACM GIS '08, November 5-7, 2008. Irvine, CA, USA

rations the spatial prepositions need to be quite flexible [6] and any model for spatial prepositions needs to take this into account.

When describing the location of an image a choice has to to be made as to which toponyms and spatial prepositions to use in the description. Due to the flexibility of the spatial prepositions and the usually large number of toponyms that could be used as the ground object in the spatial relation, in most situations there will be a number of candidate spatial prepositions and toponyms. From these candidates one spatial preposition and toponym needs to be chosen for the final caption and this choice of spatial preposition exhibits prototype effects [10], with the acceptability and frequency of use falling as the spatial relation moves away from the prototypical case. Functionality [1] [3], social use context [5] and language [4] [8] all influence this decision and when analysing spatial language it is necessary to try to classify these influences as precisely as possible to be aware of the biases this introduces into the analysis.

Mining existing data sources, such as the web, is an approach for creating models based on a much larger number of samples than is possible with human subject testing methods. Trigger phrases such as "is within walking distance of" can be used to find a number of places A and B between which the given spatial relation holds. These can be geocoded and the distance between the two calculated, which results in a set of distances that are applicable to the spatial relation mined [9]. This set of distances is then used as an input to analysis of the spatial relation "within walking distance". The danger with automatic mining of data is the inadvertent introduction of biases. In this case there is a business incentive for the website owners to describe themselves as "within walking distance" and it is unclear how this would influence the results.

3. GEOGRAPH

As an initial experiment, an analysis of an existing rich spatio-linguistic image caption data set was performed. This data set was acquired from the Geograph project[1], an open-participation project that aims to provide a representative photograph for each square kilometre of the United Kingdom and Ireland. A dump of roughly 350,000 image captions and locations forms the basis for this analysis. Due to the aim of providing representative photographs the images tend to be ground-level and panoramic for rural areas and of buildings or roads in urban areas, with captions such as "Footpath at Pirbright", "Farmland near Garthorpe" or "Lambeth Palace from Lambeth Bridge".

The project focus introduces a bias towards spatial language as it is used in image captions, but as our research interest is in spatial image caption language, this bias is actually desired. An unwanted bias is introduced by the fact that the project is built on publicly contributed data and thus skewed towards the language used by the most frequent contributors. While the skew in the original data is quite heavy (2% of the contributors provide 90% of the data), the reduction to captions using cardinal directions, and capping the number of captions per contributor that we consider, produces a good, quite large sample.

The first step in analysing the use of spatial prepositions in the Geograph data is extracting their uses from the image captions. GATE [2] is used for part of speech tagging and the identification of spatial relations. Toponym identification is handled by a simple rule. Words starting with an uppercase letter are determined to be candidate toponyms, excluding stop words such as "A" or "The". Multiple, consecutive candidate toponyms are aggregated into one multi-word candidate toponym. This rule identifies the majority of toponyms, especially toponyms for populated places. The only group of toponyms that this method does not identify are those using generic toponym classes such as "station" or "church", but there is a very limited number of such generic names in frequent use, and they can be handled by special rules.

The tagged captions are then matched against patterns of the form "<image topic> <spatial relation> <toponym>", with the optional image topic usually describing the content of the image. The hypothesis is that the GPS coordinates of the camera location, as stored in the image meta-data, and the location of the toponym matched by the pattern form one valid use of the spatial relation. Multiple uses of each spatial relation are used to build up a quantitative model.

The toponyms are geocoded using the Geonames.org service[2], which returns a point representing the centre of the toponym location. Toponym disambiguation was based on accepting only exact toponym matches and performing a simple filtering procedure. As the distances involved in the spatial language of image captions tend to be short (mostly less than 5km), a reference to the wrong toponym is immediately clear as a statistical outlier. A hard limit of 12km is enforced on the distance, filtering out the incorrect toponym disambiguations. For each of the patterns the GPS co-ordinates of the image and the location of the toponym are combined to calculate the angle and distance from the ground toponym to the image location. Combining these distance/angle pairs into sets makes it possible to create distribution plots based on a common ground location, as shown in figure 1.

The analysis is then based on these sets of distance/angle pairs. As this method combines distance and angle data from multiple captions, it is necessary to guarantee that the scale involved in all captions is the same. The area "south of" a point of interest such as a church will have a different scale to that "south of" a town or village. The Geonames.org service in addition to the location of the toponym also provides information on the toponym's type and in the data presented in this analysis only toponyms of the type *populated place* were used. As explained subsequently, this corresponds in practice to a characteristic scale of place.

3.1 Cardinal directions

3.1.1 Results

A total of 1081 data points were analysed for the cardinal directions (307 north, 330 south, 225 east and 219 west).

Distances for "north", "east" and "west" varied between 120m and 11940m, with means between 2080m and 2220m. Medians varied between 1510m and 1920m, with the interquartile ranges between 1240 and 1590m. The distances calculated for "south" varied between 40m and 11360m, with a mean of 1800m and a median of 1390m (fig. 3). The interquartile range is 1310m, with the first quartile at 820m and

[1] `http://www.geograph.org.uk`

[2] `http://www.geonames.org`

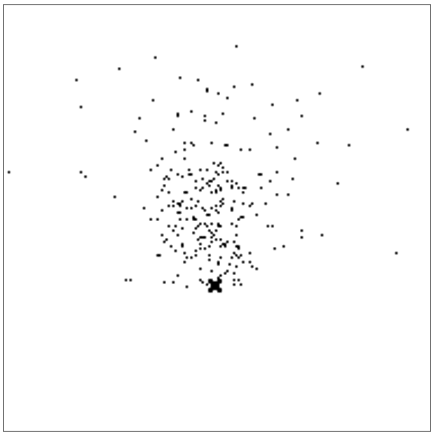

Figure 1: Figure locations for the cardinal direction "north" plotted from a common centre, representing the ground location (black X).

the third quartile at 2130m. The differences in distribution are visible in figures 2 and 3.

The north angle data shows a slight shift to the west, with a mean of 355°, whereas the south data is almost perfectly south with a mean of 181°. The east distribution is rotated by 23° towards north with a mean of 67°, while the west distribution is rotated north by 16° with a mean of 286°.

3.1.2 Discussion

The distances involved in the cardinal directions are not very high, probably because when captioning images local place names are used so that the location of the image is described as precisely as possible. Comparing the individual cardinal direction distances using ANOVAs showed that the "east" and "west" distributions can be considered basically equivalent ($p > 0.9$). While "north" shows no significant differences from "east" and "west", the median and inter-quartile distances are slightly higher indicating that with a larger sample a statistically significant difference might become apparent. The "south" distribution on the other hand shows significantly shorter distances ($p < 0.01$), when compared to the three other directions, meaning that three different models for distances of cardinal directions need to be distinguished. These patterns can also be seen in the boxplot in figure 2.

The angle data show a similar pattern, with "east" and "west" very similar. The "east" angles were mirrored across the north-south axis and Watson's two sample test applied, showing no significant differences between the "east" and "west" angle distributions. Interesting about these two distributions is that they both show statistically significant rotational shifts towards north, when compared to the same distributions centred on the east-west axis. Whether this rotational shift is caused by a subconscious error in the understanding of the east/west directions or whether it is an artefact caused by the preposition selection process, or whether some other effect is the reason cannot be determined at this point and is an open question for future research.

The extent of the ground toponym has a direct influence on the distance and angle calculated and while only populated places were analysed, populated places also cover a wide range of scales. We believe that in the case of spatial

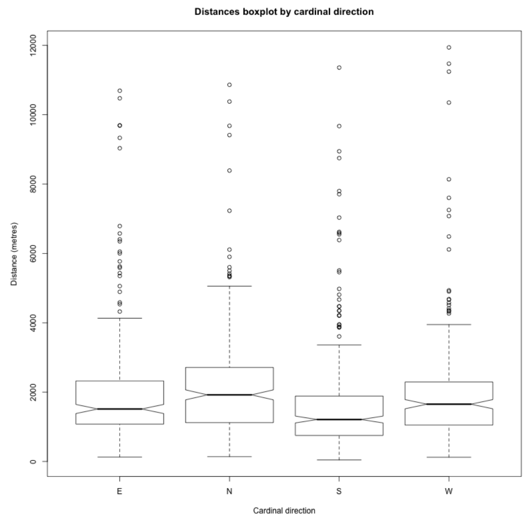

Figure 2: Boxplot of the distances by cardinal direction (east, north, south, west). Clearly visible is that "east" and "west" are very similar, while "south" is significantly different. Also visible here is that all points over about 4 to 5km are classified as outliers (empty circles).

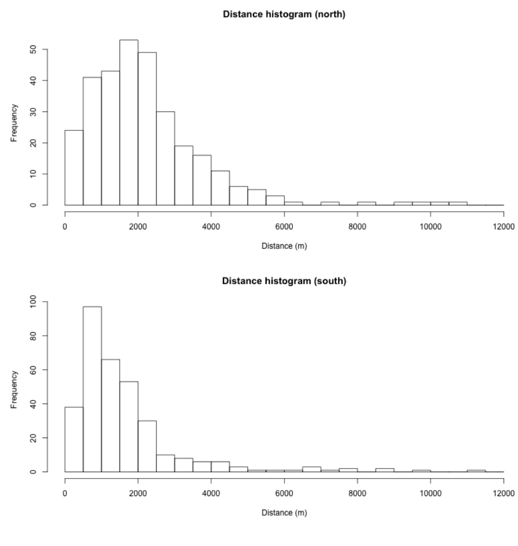

Figure 3: Distance histograms for the "north" and "south" distributions. The differences in the distribution shapes between "north" and "south" are clearly visible.

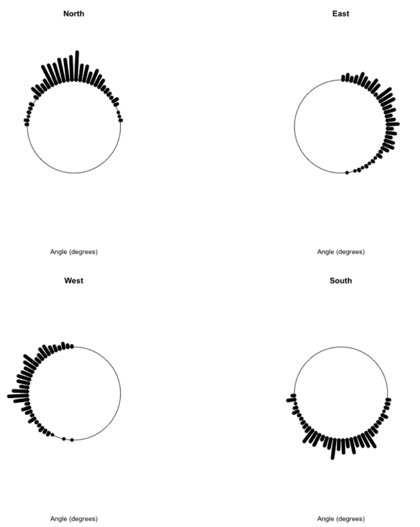

Figure 4: Circular plots for all four cardinal directions.

relations as used in image captions, the act of captioning tends to rely on very local information and this acts as a restricting force on the size of the toponyms used. This is supported by the actual distribution of distances (fig. 3) obtained and by observation of a sample of the actual ground locations used. If larger places were frequently used as the ground toponym, then to get the distribution that the data show, the distances used with the spatial preposition would have to be inversely proportional to the extent of the place. This is highly unlikely and it is much more likely that in image captions only local and small places with similar extents are used as ground toponyms.

4. CONCLUSIONS

This paper presented an experiment on the use of spatial prepositions in image captions. The experiment analysed an existing set of image captions taken from the Geograph project, determining what angles and distances are involved in the use of the cardinal directions. The main result of this analysis is that the distances involved when using cardinal directions are short, mostly less than 3000m from the ground toponym. This indicates that when captioning images very local information is used. An interesting result from this analysis is that when using “east” or “west” a rotational shift towards north is observed for the angle distributions of both cardinal directions. Possible reasons are presented, but no definitive explanation can be given at this point.

The results of this experiment can be employed in a wide variety of scenarios. Relevance ranking in the retrieval of geographic information is one scenario. New quantitative models based on the results will allow for more complex result ranking than just using distance or angle. Another scenario is geo-referencing images or documents based on their captions, if those captions use spatial prepositions. The image or document location could then be used to improve their retrieval or to add further meta-data based on the location.

The experiment was designed on the medium scale of populated places. Future work will focus on extending this research to the street and point-of-interest scale and also to further spatial prepositions. A model for representing and reasoning with the vague regions defined by the spatial prepositions is also being developed.

5. ACKNOWLEDGEMENTS

We would like to gratefully acknowledge contributors to Geograph British Isles (see `http://www.geograph.org.uk/credits/2007-02-24`, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (http://creativecommons.org/licenses/by-sa/2.5/).

This material is based upon work supported by the European Community in the TRIPOD (FP6 cr n°045335) project.

6. REFERENCES

[1] K. Coventry, A. Cangelosi, R. Rajapakse, A. Bacon, S. Newstead, D. Joyce, and L. Richards. Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. In *Spatial Cognition IV. Reasoning, Action, and Interaction*, volume 3343/2005 of *Lecture Notes in Computer Science*, pages 98–110. Springer Berlin / Heidelberg, 2005.

[2] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. Gate: A framework and graphical development environment for robust nlp tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, pages 168–175, 2002.

[3] S. Garrod, G. Ferrier, and S. Campbell. In and on: investigating the functional geometry of spatial prepositions. *Cognition*, 72(2):167–189, 1999.

[4] J. Grabowski and G. Miller. Factors affecting the use of dimensional prepositions in german and american english: Object orientation, social context, and prepositional pattern. *Journal of Psycholinguistic Research*, 29(5):517–553, 2000.

[5] J. Grabowski and P. Weiss. The prepositional inventory of languages: A factor that affects comprehension of spatial prepositions. *Language Sciences*, 18(1-2):19–35, 1996.

[6] A. Herskovits. Semantics and pragmatics of locative expressions. *Cognitive Science: A Multidisciplinary Journal*, 9(3):341–378, 1985.

[7] B. Landau and R. Jackendoff. “what” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2):217–238, 1993.

[8] S. Levinson. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: CUP, 2003.

[9] S. Schockaert, M. de Cock, and E. Kerre. Location approximation for local search services using natural language hints. *International Journal of Geographical Information Science*, 22(3):315–336, 2008.

[10] C. Vorwerg and G. Rickheit. Typicality effects in the categorization of spatial relations. In *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, volume 1404/1998 of *Lecture Notes in Computer Science*, pages 203–222. Springer Berlin / Heidelberg, 1998.