Interpreting Spatial Language in Image Captions

Noname manuscript No.
(will be inserted by the editor)

Interpreting Spatial Language in Image Captions

Mark M. Hall · Philip D. Smart · Christopher B. Jones

the date of receipt and acceptance should be inserted later

Abstract The map as a tool for accessing data has become very popular in recent years, but a lot of data does not have the necessary spatial meta-data to allow for that. Some data such as photographs however have spatial information in their captions and if this could be extracted, then they could be made available via map-based interfaces. Towards this goal we introduce a model and spatio-linguistic reasoner for interpreting the spatial information in image captions that is based upon quantitative data about spatial language use acquired directly from people. Spatial language is inherently vague and both the model and reasoner have been designed to incorporate this vagueness at the quantitative level and not only qualitatively.

Keywords spatial language · vagueness · natural language processing · spatial reasoning · field-based modelling · geographic information retrieval

1 Introduction

Photographs are inherently spatial as they are always taken at some location. This spatialness has led to the rise of photographic websites such as Geograph[1], Lockr[2] and Flickr[3] that allow their users to place their photos on a map. The problem is however that this is a rather time-consuming, manual task. To automate this process there is one source of spatial data that has so far not been utilised much and that is the image's caption.

Very often the caption will contain the image's location encoded linguistically. The problem with extracting this location is that spatial language as used in image captions is predominantely qualitative in nature and thus very vague as a small number of spatial language elements is used to describe a vast array of spatial configurations. In spatial language vague expressions such as "The statue is next to the river", "The House is near the park" or "Reading is near London" are the norm, and their interpretation does not pose any significant problem to people. The book can be "next to the glass" even if there is some distance between the two objects, the use of "near" with two objects of different scale is likewise unproblematic. Dealing with vague information is so natural to people that they are often unaware of its vague nature. At the same time a phrase such as "London is near Reading" would at least be considered odd and "London is near Cardiff" definitely wrong, indicating that while spatial language is vague and flexible, there are qualitative and quantitative constraints.

In parallel to this vague human spatial world there is the crisp computational spatial world which represents everything via points, lines and polygons. While this works sufficiently well for professional interaction with geographic information in domains such as planning, environmental analysis or geo-statistics, when it comes to providing improved access to geoinformation for lay-people the vagueness and its prevalence in natural language cannot be simply abstracted away (Altman (1994)).

In this paper we propose a system that allows for translating from vague spatial natural language into the crisp representation required for modern Geographic Information Systems. This translation is based on quantitative data on spatial language acquired using a human subject experiment (sect.

M.M. Hall · P.D. Smart · C.B. Jones
Cardiff School of Computer Science & Informatics
Cardiff University
Queen's Buildings
5 The Parade, Roath
Cardiff CF24 3AA
UK
E-mail: M.M.Hall@cs.cardiff.ac.uk, P.D.Smart@cs.cardiff.ac.uk, C.B.Jones@cs.cardiff.ac.uk

[1] `http://www.geograph.org.uk`
[2] `http://www.locr.com`
[3] `http://www.flickr.com`

3) and uses a field-based data-model that we have developed to represent vague spatial data quantitatively (sect. 4). The quantitative data and the field-based model are used in our spatio-linguistic reasoner to automatically derive the location of a spatial expression (sect. 5), the results of which have been evaluated (sect. 6) and from that evaluation a few modifications have been developed to improve the quality of the system (sect. 7).

Spatial language is incredibly flexible and to be able to handle it computationally restrictions have to be put in place as to what can be processed. Thus in this paper we will only be dealing with spatial language as it is found in image captions, which defines both a restricted sub-set of spatial language and also a characteristic scale of places in those captions.

2 Background

2.1 Spatial Language

The basic elements that are of interest to this work are objects located in space (in an image caption usually the subject of the image), the places the objects are located in or proximal to and the spatial relations between the objects and the places. This classification is based on Landau and Jackendoff (1993)'s concepts, but unlike their work there is no assumption that these concepts are linguistic universals. The concepts work in English, but no conclusion is drawn as to their applicability to any other language, as a number of studies have indicated that the concepts are not so universal after all (cmp. Kemmerer (2006); Levinson (2003); Bowerman and Choi (2003); Brown (1994)). Landau and Jackendoff also investigate object shape, but as the geo-data available at this time does not provide shape information, this will not be considered. In English the three concepts object, place, spatial relation are represented by nouns, toponyms[4] and spatial prepositions (see also Bennett and Agarwal (2007)). In the basic case an object is related to a place via a spatial preposition and in this paper the terminology of *figure* and *ground* as introduced by Talmy (1983) will be used. The *figure* is defined as the object that is located relative to the *ground* object or place and the relative location will be defined by a spatial preposition. In "statue in London" "statue" is the *figure*, while "London" is the *ground* place and the *figure* is related to the *ground* via the spatial preposition "in".

The relations are always defined relative to a reference system, of which three are usually distinguished: intrinsic, relative and absolute Levinson (2003). The intrinsic reference system defines the relationship relative to the *ground* object, the relative reference system introduces an additional object into the spatial relation that defines the reference system and the absolute reference system is based on the cardinal directions. In our work only the absolute reference system is of interest, as in captioning the locations are described at a large scale and thus only the absolute reference system provides a useful reference system.

[4] A toponym is a place with a recognisable name used in communication, where a "place" is often defined simply as a meaningful geographic location (Goodchild and Hill (2008))

2.2 Spatial Cognition and Language

The question of how cognition and language interact is not new, with views ranging from language and cognition as a single cognitive unit to language as a separate unit that builds on an underlying cognitive framework. On the side of the strongly-linked theories Sapir (1929) and Whorf et al (1956) take the view that language inherently restricts and channels how you think about something and what you can think about it, a view that is usually referred to as linguistic relativism. The opposing view of linguistic universality is usually attached to the works of Chomsky (1965) and the idea of a universal grammar that underlies all languages. Most current linguists tend to take a position somewhere in between these two extremes (see for example Tversky and Lee (1998), Mark (1989) or Mark and Frank (1995)), although that still leaves plenty of space for disagreement.

The main problem is that there is evidence to support either view. Kemmerer and Tranel (2000) use a set of experiments on two participants with brain lesions to show that depending on where the brain lesions are, it is possible for the participants to perform perceptual reasoning while failing at the linguistic spatial test and vice-versa. This indicates that spatial and language processing are at least partially separate in the brain. At the same time work by Levinson (2003) on aboriginal languages in Australia and South America shows that some languages provide only absolute systems of reference. Such languages would use "the man is standing north of the house" where in English one might say "the man is standing in front of the house" or "the man is standing left of the house" depending on the speakers position. Levinson reasons that if the language only provides the necessary words to encode absolute relations then the brain's spatial cognition system must keep track of all objects locations in an overview-map-like representation, indicating that language has a direct influence on spatial cognition. One could say that it is the environment that influences rather than language, an approach taken by Li and Gleitman (2002), although Levinson et al (2002) provide further experimental data indicating that the environment's influence in the Li and Gleitman experiments was caused by the experimental setup not testing the relevant brain functions. At the same time Mark et al (2007) analyse a different aboriginal language that primarily uses topographic features for

Fig. 1 The grey apple is "in" the bowl even though it is not contained in the area of the bowl itself.

classification, while functional aspects such as seasonal water is classified using an event-based formalism (flood, bit of water). More evidence of a language influence on spatial reasoning is provided by Klippel and Montello (2007) with a study on direction evaluations. Their experiments showed that if participants were told that they would have to label the directions they were evaluating, then that influenced their judgements. The interesting thing is that the results indicate that language's influence on spatial reasoning can be switched on or off depending on the context.

It seems to be that there is interaction between language and spatial cognition, but at the same time there is also a cognitive separation between the two. In this paper a utilitarian approach is taken that acknowledges that this discussion exists and that it implicitly influences the design of a computational system that handles spatial natural language, but the central idea is to teach the system to understand natural language in a human similar way, irrespective of whether the methods chosen parallel the human "implementation".The focus is on results that are similar to what people would produce, not copying human mental models and methods.

2.3 Modelling Spatial Language

Spatial language needs to be represented computationally and as spatial prepositions represent spatial relations linguistically the first choice would seem to be a first-oder logic based approach. The problem with that however is that spatial prepositions are so flexible and when and how they can be used depends on so many factors (context, function, ...) that a logic-based approach will either reject many valid situations ore accept situations that humans would judge to be incorrect (see Herskovits (1986), and the work by Coventry et al (2001) and Garrod et al (1999)). It seems to be that spatial prepositions exhibit prototype-effects, where there is a default interpretation and constraints imposed by this default interpretation can be relaxed if there is contextual information that allows for that (see Vorwerg and Rickheit (1998)). Figure 1 shows an example of a non-default containment relation, where the top-most apple is "in" the bowl, because the functional link with the bowl (if you move the bowl, the apple also moves) overrides the strict containment constraint.

Gärdenfors (2000) "conceptual spaces" are another good computational model, but they require that all aspects of the reasoning process are performed quantitatively. An alternative is to treat spatial language as a set of instructions as proposed by Miller and Johnson-Laird (1976). Thus a spatial expression "rocks near Stackpole Head" is translated into a procedural representation such as "near(rocks, Stackpole Head)". A more complex expression is treated as a nested set of such procedural statements. The statements can then be executed and in the processing of each procedure both qualitative and quantitative reasoning steps can be combined, allowing it to support more qualitative prototype effects and also purely quantitative effects such as distance or angle.

As a final possible model, because it influenced some early thought in the development of this work, for representing spatial language is the concept of "image schemata". Developed by Lakoff and Johnson (1980) it postulates that since all humans live in the same world, governed by the same physical constraints there is a set of universal concepts that underpin reasoning. Since most of these concepts are spatial in nature (path, surface, container, ...; see Johnson (1987) for a full list) they have been used to underpin a variety of GIS research, such as spatial algebras for maps and room-space (Couclelis and Gottsegen (1997)), interoperability (Frank and Raubal (1999)), wayfinding (Raubal and Worboys (1999)), document retrieval (Fabrikant and Buttenfield (2001), and basic GIS concepts (Kuhn (2002); Mark (1989)). The problem is that the original theory makes very strong claims as to image schemata being the actual representation in people's minds, which is based only on English language examples and very little empirical data to support the claims. Thus while some of the concepts have been retained as they provide easily understandable structures, the assumption of universality and actual use as a mental model have been discarded in our work.

2.4 Vagueness

As has been touched upon in the introduction, vagueness is a basic aspect of spatial reality (see Fisher (2000); Parsons (1996)). While man-made objects such as buildings, districts, countries have clearly defined boundaries, natural features such as coastlines, forests, mountains tend to have vague boundaries (Smith and Varzi (1997)). Historically, as map-making was done by hand, the integration of vague information into the map was performed via the use of symbol density, shading, colours, or any other method the map-maker wanted to employ. The transition to computer-based map-making reduced maps to the crisp representations favoured by digital systems. Thus in Geographic Information Science (GIS) research the early models for reasoning about space were also crisp (see Randell et al (1992); Egenhofer (1991); Güting and Schneider (1993)).

That was sufficient for the initial on qualitative reasoning and topological relations, but the problem of vague ar-

eas remained and extensions to the basic crisp models were proposed by Cohn and Gotts (1996b); Clementini and Felice (1997); Schneider (1996). These models replaced the original crisp boundaries with an extended boundary that represented the transition between the inner area where the phenomenon being described certainly applied and the outer area where the phenomenon did not apply at all. The advantage of this was that it allowed for qualitative reasoning with vague spatial information, without having to define how in detail how the vagueness worked.

Another method for representing vague information that is similar to the broad-boundary models uses rough-sets as the underlying representation (see Ahlqvist et al (1998) or Bittner and Stell (2003)). A rough-set consists of a pair of crisp sets, one representing the lower-bound and one the upper-bound. Those elements that are in the upper-bound set, but not in the lower-bound set basically define the elements of the broad-boundary in the broad-boundary models.

2.4.1 Fuzzy models for vagueness

Although these models provide a way of dealing with vagueness, they are still qualitative and do not specify how the transition in the broad boundary works, whether it is linear, follows an exponential decay curve or in the case of a partially crisp boundaries contains discontinuities. To enable such a quantitative representation it is necessary to move to a fuzzy-set (Zadeh (1965)) based approach[5]. Fuzzy sets use a membership function to represent to what degree an object is part of the set. Usually the range $[0,1]$ is used, with 0 representing classical non-membership and 1 classical membership. In between these two extremes, the membership value varies according to a pre-defined membership function and it is the definition of this membership function that is one of the more difficult aspects of fuzzy-approaches, as a balance has to be found between a simpler membership function abstracted from the source data and a complex membership function that is directly derived from the phenomenon. Both approaches have their advantages, as the simpler membership function tend to have better understood behaviour when used with fuzzy operators, while the directly derived function provides a better representation of the vague phenomenon, the decision between the two approaches has to be made on a problem-by-problem basis (see Robinson (2003)).

A number of fuzzy models have been devised for dealing with topological reasoning (Schneider (2001); Winter (2000)), geomorphological reasoning (Fisher et al (2004)), general spatial representations (Hwang and Thill (2005); Tang (2004); Pfoser et al (2005); Wang and Hall (1996)) and most importantly linguistic reasoning (Schockaert et al (2008); Gapp (1994); Robinson (2000); Worboys (2001); Worboys et al (2004); Gahegan (1995); Fuhr et al (1995)). The work by Schockaert et al (2008) illustrates that by using a focused corpus (in their case of hotel websites) it is possible to derive a fuzzy representation of an arbitrary spatial phrase such as "within walking distance", an approach that was used to derive an initial overview of how spatial prepositions were used in image captions (Hall and Jones (2008)). The Worboys and Robinson experiments and the work by Fisher and Orf (1991), although they did not use fuzzy modelling, showed that it is possible to acquire fuzzy representations directly from people using a map like approach. An interesting aspect is that they describe varying levels of success in their experiments. While Worboys and Robinson report that they have created multi-person fuzzy models for "near" at their respective scales (Worboys at the campus scale, Robinson at the inter-town scale), Fisher and Orf state that the results of their experiment do not allow them to create a formal representation of "near" (again at the campus scale). The results presented in this paper show a picture that is closer to the Worboys and Robinson results, in that the results are sufficiently similar to allow their use in spatial reasoning, but similar to Fisher and Orf they also show a lot of variation between people's interpretations.

A side-effect of using a fuzzy-based approach is that, as Altman (1994) illustrates, the detailed representation of the vague phenomenon can be retained throughout the spatial reasoning process and only at the final decision point is the fuzzy representation reduced to a true/false decision, a decision that with qualitative representations has to be taken much earlier in the reasoning process.

2.4.2 Field-based models for vagueness

The various fuzzy representations are very powerful tools for representing vague knowledge, but to use them it is always necessary to represent the vague knowledge as one or more membership functions and in the case of multiple membership functions the methods for combining them needs to be specified, which as mentioned above can be quite difficult. To avoid this difficulty the vague data can be represented using a field model, which maps coordinates to values (see Liu et al (2008); Laurini and Pariente (1996)). Fields represent geographical knowledge as a continuum of values, contrasting with object-based representations that split geographic space up into a set of more or less clearly delineated objects (see Couclelis (1992); Goodchild (1992)). As an example a field model could use a matrix of water depth values to represent a lake, while an object model would use a polygon to define the lake's boundary by fiat. However, the

[5] There is also the possibility of using an approach based on supervaluation (see Bennet (2001) or Kulik (2001)), which allows for the creation of a set of boundaries that describe the gradual transition from the definite to the definitely-not area, similar to iso-lines used to represent height on conventional maps. The fuzzy methodology seems to be more frequently used, thus is given prominence.

two concepts are not in opposition to each other, there are cases where either model can be used (or an intermediate model such as the one by Erwig and Schneider (1997)) and also cases where one of the two is more suited. Fields tend to be better suited to representing spatial information, while an object model makes the manipulation of the individual spatial objects easier.

Fields have been used in GIS to determine the precise location of an object in a scene description (see Yamada et al (1992)), for locating areas described in biological specimen records (see Wieczorek et al (2004); Liu et al (2009)) and for robot navigation (see Mukerjee et al (2000)). In this paper we use a similar field-based approach, but where our approach differs is that it is based on quantitative data about vague language acquired directly from people and that it treats vagueness as an inherent property of spatial language and not as uncertainty or error.

3 Linguistic data-acquisition

3.1 Structural caption analysis experiment

To develop an initial understanding of the quantitative aspects of spatial language as used in image captions, an existing large geocoded image caption data set was analysed. This data set was acquired from the Geograph project[6], 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 Geograph data was analysed along two separate lines. Quantitative data on how the spatial prepositions "near" and the cardinal directions are used, the results of which can be found in Hall and Jones (2008), and a qualitative analysis of the syntactical patterns found in image captions.

The quantitative analysis in Hall and Jones (2008) looked for simple patterns of the form "<subject> <spatial relation> <toponym>" and used those patterns to extract the spatial information of interest for the quantitative analysis. For the caption interpretation algorithms presented later in this work, a more complete understanding of the structures involved in image captions is required. To this end a syntactical and statistical analysis of image caption language has been undertaken using collocation.

While a pure word collocation analysis in image captions can produce interesting results, the problem is that there is so much variation in how place-names are structured and how they are combined with adjectives and syntactic elements such as determiners. This variation amounts to noise that obscures the structural pattern signals. To avoid this the statistical analysis is based on a syntactically tagged part of speach (POS) representation of image captions and not directly on the image captions.

Tag 1	Tag 2	Frequency	Generalisation
NNP	NNP	629	NPhr
IN	NNP	146	IPhr
NNP	,	135	CommaPhr
,	NNP	131	CommaPhr
NNP	IN	97	
NNP	NN	69	NPhr
DET	NNP	68	NPhr
NN	IN	58	

Table 1 Top eight collocations derived from 580 caption (out of a set of ∼350000 captions, but only one caption per user considered). Also shown is the POS tag that the individual collocations will be generalised as. NNP - Proper noun, IN - Preposition, DET - Determiner, D NPhr - Noun phrase, IPhr - Prepositional phrase, CommaPhr - Comma separated phrase.

Calculating the POS tag collocations produces the collocation distribution[7] shown in table 1. As can be seen in that table, the most frequent collocations are noun - noun combinations, preposition - noun patterns and noun - comma patterns. The POS tagging has reduced but not eliminated the amount of variation. To further reduce this variation the most frequent collocations are used as generalisation rules. These combine various noun combinations into noun-phrases (lines 1,6,7), create prepositional phrases (line 2) and comma separated phrases (lines 3,4). After the generalisation rules are defined the process starts from the beginning, except this time the generalisation rules are applied before the collocations are calculated.

The generalisation rules are applied from right to left, as in English the more specific elements of a phrase are towards the front. Otherwise if the rules were applied from left to right then a caption such as "The old gate house" would result in the syntactical tree structure shown in figure 2. This structure would incorrectly imply that the adjective (JJ) "old" belongs to the noun (NN) "gate" instead of the combined noun "gate house". Similarly the determiner (DET) refers to the whole noun phrase (NPhr) and not just the "old gate". Applying the rules from right to left produces the correct syntactical tree (fig. 3).

This process of generating generalisation rules from the collocations and then recalculating the collocations is repeated until the remaining collocations form only a tiny part

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

[7] To avoid a bias being introduced by a small group of frequent contributors producing most of the captions only one caption per contributor was considered. This reduces the number of captions from around 350000 to 580.

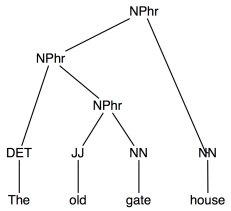

Fig. 2 Incorrect syntactical tree structure generated by applying the rules from left to right.

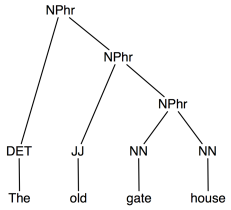

Fig. 3 Correct syntactical tree structure generated by applying the rules from right to left.

Tag	Frequency	Example
NPhr	242	"Merthy Tydfil"
FigureGroundPhrase	131	"Sheep near Stackpole Head"
ContainmentPhrase	96	"Roath Park, Cardiff, Wales"

Table 2 Top 3 full-caption patterns derived from 580 caption (out of a set of ∼350000 captions, but only one caption per user considered)

of the caption set ($< 0.5\%$) and for all other captions the generalisation rules create a syntactical tree for the whole caption (tab. 2).

The generalisation process identified three major caption patterns. These are captions that consist only of noun phrases (fig. 4), captions consisting of a noun phrase plus a prepositional phrase and captions consisting of a list of comma-separated noun phrases. In terms of content these equate to captions consisting only of a place name ("Merthyr Tydfil"), captions where something is related to a place name via a spatial preposition ("Sheep near Stackpole Head") and a list of place names specifying a containment hierarchy ("Roath Park, Cardiff, Wales"). The collocation analysis clearly shows that the linguistic patterns found in image captions are quite simple, probably due to the act of captioning being time-consuming and thus the simplest possible caption that conveys the necessary information is chosen (the three caption types identified in table 2 represent about 80% of all captions. Knowledge of the caption patterns described here is used in the caption interpretation reasoner to determine how and where to extract spatial information from image captions.

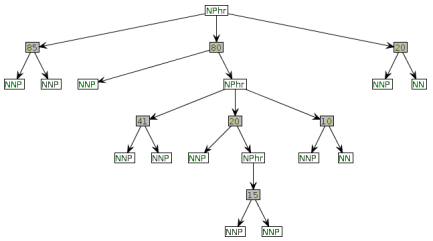

Fig. 4 Top three NPhr (noun phrase) structures found in the data. POS tags in the white boxes, collocation frequencies in the grey boxes. A POS tag with multiple collocation frequencies indicates the collocations that have been generalised into that POS tag and their frequency. The diagram shows nicely that most NPhr consist of two or more NNPs (proper nouns).

3.2 Quantitative data acquisition experiment

The Geograph analysis in Hall and Jones (2008) provided an initial view onto the quantitative aspects of spatial language, but to implement a spatial language interpreter a cleaner data-set on the use of spatial prepositions was required. To create such a data-set a human-subject experiment was designed based on the Geograph analysis results. A total of 24 participants were recruited from undergraduates and staff at Cardiff University.

The experiment itself consisted of eight questions with each question consisting of a map showing the test area, a primer phrase and the answer section. The primer phrase used was "This photo was taken in _____ which is <spatial preposition> Cowbridge", with each question testing a different spatial preposition. The participants were instructed to rate how applicable the primer phrase was for each of the test places shown on the map using a 9-point Likert-type scale. On the rating scale 1 indicated that the primer phrase did not apply at all, while a rating of 9 indicated perfect applicability. To avoid introducing unwanted biases the places on the map were shown as points (to avoid relative size effects as in Morrow and Clark (1988)), the places were distributed on the map to provide even coverage without suggesting any regularity and in the answer section the places were listed alphabetically. Results are interpreted based on their median values and inter-quartile range. Inter-participant agreement was defined as high if the inter-quartile range was 0 or 1, as medium if the inter-quartile range was 2 and low for any higher inter-quartile ranges.

3.2.1 Near

Table 3 shows median and inter-quartile range for all toponyms in the question (fig. 5). Two toponyms have high inter-participant agreement, seven toponyms have medium agreement and seven toponyms have low agreement. Median values can also be split into three groups, an inner ring with medians of 8 or 9, a group with medians of 5 and a distant group with medians of 3 or lower.

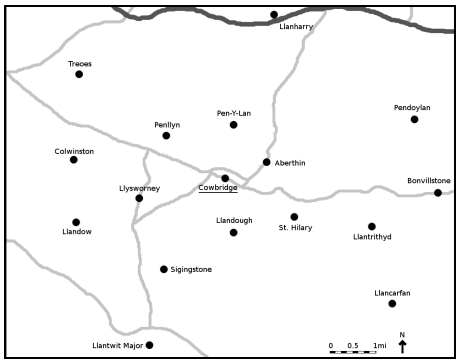

Fig. 5 Map displayed for the "near" and cardinal direction questions, with "Cowbridge" as the *ground* location.

	Near	North	East	South	West
Aberthin	9 / 0.25	6 / 1.5	8 / 1	1 / 0	1 / 0
Bonvillstone	3 / 3.5	1 / 1.5	9 / 0	1 / 2	1 / 0
Colwinston	5 / 3	1 / 2.5	1 / 0	1 / 0	9 / 1
Llancarfan	2 / 2.5	1 / 0	7 / 2	5 / 6	1 / 0
Llandough	8 / 1	1 / 0	2 / 3.5	9 / 0	1 / 2
Llandow	5.5 / 3	1 / 0.5	1 / 0	5 / 4	8 / 1
Llanharry	2 / 2.5	8.5 / 1	2 / 3.5	1 / 0	1 / 0
Llanthrithyd	5 /3	1 / 0	8 / 2	5 / 2.25	1 / 0
Llantwit Major	2 / 2.5	1 / 0	1 / 0	8 / 1	3 / 3.5
Llysworney	8 / 2	1 / 1	1 / 0	3 / 3.25	9 / 0
Pendoylan	3 / 3	5 / 3	7 / 2	1 / 0	1 / 0
Penllyn	7.5 / 2	6 / 2	1 / 0	1 / 0	7 / 2
Pen-Y-Lan	8 / 2	9 / 0	2 / 3	1 / 0	1 / 1
Sigingstone	5 / 3	1 / 0	1 / 0	8 / 1	5.5 / 3
St Hilary	8 / 3	1 / 0.5	/ 1	7 / 1.5	1/ 0
Treoes	3 / 2	5 / 2	1 / 0	1 / 0	7 / 2

Table 3 Median values and inter-quartile range per toponym for all questions. Values are formatted median / inter-quartile range.

Statistical significance calculated via chi-square tests shows significance for 14 of the 16 toponyms at $p < 0.05$, with the exception of "Bonvillstone" (p = 0.24) and "Pendoylan" (p = 0.12), indicating that the participants were not just randomly selecting an rating. The results can thus be used as valid inputs into the analysis and model of "near".

Discussion The central result from the data is that the primary factor when deciding whether a place is near another place is distance. This creates a kind of banding effect, an inner circle with high median values, a middle circle and then an outer circle with low median values. To verify this a linear model $applicability = a + b \cdot distance$ has been fitted to the median and distance values, with the parameters fitted at $a = 10.595191$, $b = -0.001164$, both parameters significant at $p < 0.001$. The linear model also fits quite well with the fuzzy model provided by Worboys (2001) for "near" on the campus scale, which also uses a linear membership function.

Inter-participant agreement is not very high for this question. Only two toponyms "Aberthin" and "Llandough" have a high inter-participant agreement. All other toponyms have either medium or low inter-participant agreement. This indicates that while distance is the primary factor, there is little agreement between the participants on how the applicability for "near" varies with distance.

Road connectivity does not seem to have any statistically significant influence on the applicability ratings. The answers for "Llysworney" and "St Hilary" were compared using a chi-square test. Both are almost the same distance from Cowbridge and at a similar angle, but "Llysworney" was shown as connected, while "St Hilary" was not. The chi-square test did not show significance with p = 0.4657 ($\chi^2 = 5.6313$, df = 6). To verify that road connectivity does not have an effect over longer distances, "Bonvillstone" (connected) and "Llancarfan" (unconnected) were similarly compared, and also showed no significant differences with p = 0.6061 ($\chi^2 = 5.4429$, df = 7).

To test whether the participants were simply applying diagrammatic reasoning on the map the results were tested using a number of diagrammatic models (broad-boundary, distance from centre, relative distance from centre, distance in a voronoi diagram). None of the models tested showed high agreement with the results from the experiment, from which we conclude that while diagrammatic reasoning might influence our results, the lack of agreement to the models tested indicates that the results can be used as a model for the applicability of "near".

3.2.2 Cardinal directions

Table 3 shows median and inter-quartile range for all toponyms (fig. 5). The table clearly shows that inter-participant agreement is high for those toponyms that are outside the half-planes defined by the cardinal directions. The median value for all toponyms in this group is 1 and the inter-quartile range at 1 or lower, indicating high inter-participant agreement. The same is true for those toponyms that lie closest to the prototypical axis for the cardinal direction, which have medians of 8 or 9 and again the inter-quartile range is 1 or lower. The toponyms that lie between these two extremes have median values that decrease with an increasing angle from the prototypical axis, with a corresponding increase in the inter-quartile range.

Statistical significance calculated via chi-square tests shows significance for almost all toponyms at $p < 0.05$. The exceptions are "Treoes" (p = 0.15) in the "north" data, "Llancarfan" (p=0.18), "Llandow" (p = 0.34), "Llantrithyd" (p = 0.15) and "Llysworney" (p = 0.40) in the "south" data, "Llantwit Major" (p = 0.18) and "Pen-Y-Lan" (p = 0.21) for the "west" data. The exceptions are all borderline cases for the respective cardinal directions, thus the lack of statistical significance only indicates that there people use very differing mental models for the cardinal directions.

Discussion For the cardinal directions the main factor in determining applicability is the angle from the prototypical axis for each cardinal direction. A distinct banding effect can be seen in the data, the first band with angles +/-45° of the prototypical axis having very high median values and also a high inter-participant agreement. The third band contains those toponyms outside the cardinal direction's half-plane, which have low median values and high inter-participant agreement. Between those two bands lies the +/-45° to +/-90° band which has low inter-participant agreement and where the median values decrease towards 1 as the angle increases.

The importance of distance is not quite clear. In the "south of" data "St Hilary" and "Llancarfan" have almost exactly the same angle, but "Llancarfan" is more than twice the distance. While the median for "Llancarfan" is lower than for "St Hilary", a chi-square test shows no statistically significant differences between the two distributions with p = 0.3410 (χ^2 = 9.0159, df = 8). On the other hand in the "north of" and "west of" data, for the pairs "Aberthin" / "Pendoylan" and "Llysworney" / "Llandow" the more distant toponym has a lower median and chi-square tests show significant differences ($p < 0.05$ and $p < 0.01$ respectively). Contrary to the "north" data, the "east" data shows no statistically significant differences between "Aberthin" and "Pendoylan". This indicates that the effect of distance is weak and only relevant when two toponyms with almost equal angles have to be compared.

Providing an insight into how strongly local knowledge influences the rating results, two participants who rated their knowledge of the area as 8 or higher, had ratings higher than 1 for "Colwinston", "Penllyn", "Pen-Y-Lan" and "Aberthin" in the "south" data. While the map shows all places as points, in reality "Cowbridge" has a larger extent with a bulge towards the north, so these places are actually slightly south of parts of "Cowbridge". These two participants were either relying primarily on their own knowledge or at least overriding the map where they had more detailed knowledge. How this local knowledge should be integrated into computational models remains to be investigated.

The cardinal direction results have been compared to angle-based models and none of the models tested provided good predictions, thus we can conclude that the cardinal direction data also represents more than pure diagrammatic reasoning.

4 Vague Field Model

Computers are not designed to represent vague information as is acquired from the experiment described in the previous section. Their grounding in binary logic forces the reduction of everything that is to be represented to a binary form. The result of this is that to represent vague information a translation has to be made between the vague data and the binary crisp representation of the underlying computational system. Initial approaches to representing vague spatial information computationally were the broad-boundary models (Cohn and Gotts (1996a); Clementini and Felice (1996)). Later fuzzy models were proposed as representations of vague spatial information (Altman (1994); Fisher (2000); Robinson (2000); Schneider (2000)) in order to overcome the simplifications of the broad-boundary models and create a more realistic model of the vague spatial information, but bring with them the problem of how to define the fuzzy membership function (Robinson (2003)).

4.1 Definition

Instead of using a functional fuzzy representation we define a field-based model (cmp. Couclelis (1992); Goodchild (1992); Liu et al (2008)) for representing vague phenomena. Conceptually the vague-field is a two-dimensional, unbounded, continuous, scalar field defined on a projected, euclidean coordinate system[8], with the field's values in the range $[0, 1]$. Representing vagueness directly is complicated, as there is no ready scale for vagueness. Instead the field represents the applicability of a given vague spatial expression relative to a *ground* location. A field value of 0 is interpreted as "relative to the *ground* location the spatial expression does not apply at all at this point", a value of 1 as "this point is a perfect example of the spatial expression relative to the *ground*" and the values in between represent partial applicability of a varying degree. The applicability value acts as a surrogate for the vagueness, as it is derived directly from the underlying vagueness and the gradual change of applicability is caused by the vague nature of the spatial expression. It is important to re-iterate that the applicability values represent vagueness and not uncertainty or error (cmp. Guo et al (2008); Wieczorek et al (2004)).

Representing vague information using this vague-field structure has the advantage of that the field representation is detached from the method used to acquire and model the vague phenomenon. It is thus possible to populate vague-fields based on point clouds, interpolation or functional representations, although the fields used in this paper are all based on interpolation of the data acquired in the experiment described in section 3.2.

4.2 Operations

On this vague-field structure four operations are defined to `instantiate`, `read`, `combine`, and `crisp` the vague-field.

[8] See Appendix A for the technical aspects of the vague-field

Instantiation The `instantiation` operation takes the source data from which the field is defined and transforms it into the continuous representation needed by the vague-field and can be performed using any kind of density estimation, interpolation or functional method. The only requirement is that it produces a continuous, two-dimensional field of scalar values. In this paper interpolation using ordinary kriging (Krige (1951); Matheron (1962)) is used, however instantiation using Kernel Density Estimation or based on functional representations is also possible, depending on what type of source data is available (cmp. Liu et al (2010)).

Accessing the field values The `read` operation provides access to the vague-field's applicability values. As the conceptual structure of an unbounded, continuous scalar field is not technically implementable, the `read` operation translates between the unbounded, continuous field representation and the internal, bounded, discrete structure. This guarantees that all algorithms that use the vague-field can treat the fields as unbounded and continuous, while at the same time allowing the vague-field to be implemented computationally.

Combining fields In some cases a spatial expression contains more than one spatial preposition, for example "Fields near High Binton in Essex", leading to the situation where two or more vague-fields need to be combined. To support this a simple combination function (eq. 1) is defined on the vague-field, in which the combined field value cf is the sum of the source field's f_i values normalised to the $[0,1]$ range. The normalisation factor nf is defined as the maximum of the combined field's values[9] and guarantees that the values at the combined field's maxima are 1.

$$cf(x,y) = \frac{\sum_{i=0}^{n} f_i(x,y)}{nf} \tag{1}$$

The normalisation is necessary to ensure that the semantics of the vague-field remain unchanged and that at those points where the combined field has its maxima, the field has a value of 1 indicating that "this point is a perfect example of the spatial expression relative to the *ground* location".

Crisping the vague-field The vague-field model is a very powerful representation for vague phenomena and in the next section a reasoner for directly using vague-fields to interpret spatial language is described. At the same time to facilitate the integration of vague-field algorithms with existing GI systems and algorithms it is necessary to map the vague representation into the crisp structures (points, lines, polygons) that are prevalent in current GI systems and algorithms.

[9] See Appendix A for details on how it is calculated

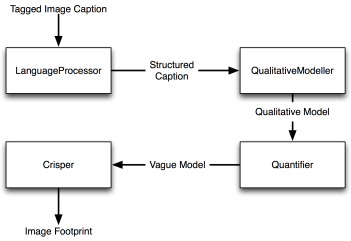

Fig. 6 The components that make up the caption interpretation system and the data elements that are passed between the components.

To support this mapping an active-contour based crisping algorithm for transforming the vague-field representation into a crisp polygon has been developed. There are other, existing methods for this process, such as α-cuts (see Klir and Yuan (1995); Purves et al (2005)) or centre-of-area methods (see Power et al (2001)), however using active-contours makes it possible to produce smoother, continuous polygons and also to easily integrate further external influences such as shorelines, mountains, or country boundaries. Figure 30 shows how the active contour is used to determine a crisp polygon for a vague-field.

5 Interpreting Spatial Expressions

Using the spatial language data and the vague-field model a spatio-linguistic reasoner for automatically interpreting the spatial language in image captions has been developed. Similar to the work of Friedman et al (2001) and Srihari and Rapaport (1990) the system uses domain-specific heuristics to identify and extract the spatial information from image captions. The reasoner consists of four components that are chained in a linear fashion (fig. 6). The `LanguageProcessor` takes the image caption and applies a sub-set of the generalisation rules identified in section 3.1 to it in order to create a more generic and regular structure out of the flexible raw caption syntax. This generalised caption is then used by the `QualitativeModeller` which extracts the spatial information and builds a qualitative model representing the photo's location. The `Quantifier` enriches the qualitative model with quantitative data transforming it into a vague-field model that quantitatively describes the likely location of the photograph. From the vague-field model the `Crisper` creates a polygon representing the image's likely location, which can then be used to further process the likely image location in existing GI systems.

This crisp polygon of the image's likely location will be referred to as the image's footprint. It represents the area in which the spatio-linguistic reasoner believes the photograph is likely to have been taken, and it is important to note that it is a belief, there is no proof that the photograph was actually taken within the delineated area. This is not a failure

in the reasoner, but an instrinsic property of linguistic information transfer, as Hall (1980) argues that "decodings do not follow inevitably from encodings". Assuming speaker A linguistically encodes information I_A into the message M using an encoding e_A and assuming that there is no loss or corruption of the message during the transfer to the listener B, then B will decode M into the received information I_B using a decoding d_B In an ideal world d_A is the inverse of e_A and thus $I_A = I_B$. The problem with this however is that what is transferred is only the message M and no information on how e_A was constructed. B is thus forced to create d_B based what they believe A meant and on their own contextual knowledge[10]. In the same way when an image is captioned the spatial information is encoded linguistically and the same information cannot be decoded exactly[11], because decoding the caption is always performed within the decoder's context, experience and intent. Thus the spatio-linguistic reasoner simply becomes another decoder, albeit with a simpler spatio-linguistic model and less contextual knowledge. The experimental data forms its experience and knowledge of the quantitative aspects of spatial linguistics, while the algorithms presented in this section represent its thought structures.

The reasoner is only able to interpret image captions that deal with places specified at the regional scale of towns and villages, mainly for practical reasons. At the time the work in the project was begun, there was very little geo-data available that could be used to geocode toponyms at the intra-city scale such as roads or points-of-interest. Additionally the number of ambiguous toponyms is even higher at the city scale making it much harder to correctly disambiguate the toponyms given in an image caption in those cases where the necessary geo-data would have been available.

5.1 Image caption pre-processing

The image caption is pre-processed before passing into the core spatial language interpreter.

5.1.1 Part-of-speech tagging

The first step in pre-processing the data is to annotate the caption with syntactic information by adding part-of-speech (POS) tags to each word. There are a number of POS-taggers available and in the TRIPOD project the GATE tagger developed at Sheffield University was chosen (Cunningham et al (2002)), as they were one of the project partners. The caption is passed to the GATE tagger as a list of words and the tagger then assigns POS tags to each word (tab. 4).

	NN	IN	NNP	NN
Caption	Sheep	near	Stackpole	Head

Table 4 The caption "Sheep near Stackpole Head" and its POS tags.

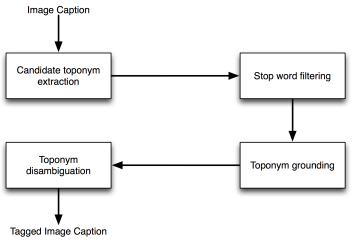

Fig. 7 The steps the caption geocoding process runs through.

5.1.2 Caption geocoding

The second pre-processing step is that of toponym resolution (Leidner (2007)), which is a special type of named entity recognition[12] (NER) with the goal of identifying noun phrases that are placenames (Ravin and Wacholder (1996)). Additionally the identified toponyms must then be assigned coordinates in a process called toponym grounding. Processing the caption to identify the toponyms is done in four steps (fig. 7). First candidate named entities and concepts are extracted from the caption, these candidate named entities are then filtered for stop-words and concepts, the filtered named entities are grounded by looking them up in a meta-gazetteer (Smart et al (2010)) and finally the grounding which may be ambiguous is disambiguated. An overview of each step will now be given.

Candidate toponym identification To extract the candidate named entities heuristics are used to extract three candidate toponym types. Unary Candidate Toponyms (UCT) are continuous sets of one or more proper nouns (NNP) that identify a candidate toponym (eq. 2). Binary Candidate Toponyms (BCT) are two UCTs that are linked by a preposition (eq. 3), while Transitive Candidate Toponyms (TCT) are lists of at least two UCTs that are either all linked by the preposition "in" or by commas and they represent a containment hierarchy (eq. 4). All possible combinations of BCTs and TCTs are generated in addition to the UCTs to provide the disambiguation processing with as much contextual information as possible (tab. 5).

$$\mathrm{UCT} = \mathrm{NP}^{+} \tag{2}$$

[10] As Lodge (1984) phrases it "It's not so easy, every decoding is a new encoding".

[11] This can be seen in the evaluation results in section 6 where the low inter-participant agreement indicates the large number of different possible encodings and decodings.

[12] Named Entity Recognition is the task of identifying noun phrases that refer to specific individuals whether these be people, companies, dates,...

NNP	*IN*	NNP	*NNP*	*IN*	*DET*	*NNP*	*NNP*
Pontsticill	near	Merthyr	Tydfil	in	the	Brecon	Beacons
UCTs	Pontsticill, Merthyr Tydfil, Brecon Beacons						
BCTs	Pontsticill + Merthyr Tydfil						
TCTs	Merthyr Tydfil + Brecon Beacons						

Table 5 An example caption with its POS tagging and the UCTs, BCTs and TCTs extracted from the POS tagged caption.

$$\text{BCT} = \text{UCT (IN} \mid \text{COMMA) UCT} \quad (3)$$

$$\text{TCT} = \text{UCT ((“in” } \mid \text{ COMMA) UCT)}^{+} \quad (4)$$

When identifying UCTs the algorithm also checks whether the first non-NNP word is a concept from the Concept Ontology (an ontology about photographic concepts developed by Edwardes and Purves (2007) for the Tripod project). If it is then that information is added to the UCT to help with the disambiguation process.

Candidate toponym filtering The candidate toponym identification process creates a large set of candidate toponyms based mostly on syntactical aspects, blindly creating candidate toponyms for anything that is identified as a proper noun. The filtering step removes all candidate toponyms that are identified as concepts through a lookup in the ConceptOntology and also those that are stop-words (“The”, “An”, ...) to create a slightly cleaner list of candidate toponyms.

Initial toponym grounding The unary, binary and transitive candidate toponyms are then grounded by looking them up in the geo-data sources that the ToponymOntology manages. In an initial step the unary candidate toponyms and the first candidate toponym in the binary and transitive candidate toponyms are grounded by querying the geo-data. All candidate toponyms for which exactly one toponym in the datasources exists are defined as unambiguous and added to the set of geocoded toponyms. Candidate toponyms for which no entries exist in the geo-data are removed from the list of toponyms as they cannot be geocoded. The remaining candidate toponyms for which more than one entry exists in the geo-data are marked as ambiguous and retained for disambiguation in the next step.

Toponym disambiguation The Toponym disambiguation steps are similar to other custom heuristic and spatial correlation approaches shown in the literature (Smith and Crane (2001); Andogah et al (2008); Buscaldi and Rosso (2008)). Initial disambiguation works using the BCTs and TCTs. The candidate locations of the first UCT in the BCT or TCT is used to query Yahoo’s WhereOnEarth (WOE) service[13], which returns a containment hierarchy. As the BCTs and TCTs also define containment hierarchies they are compared with the WOE hierarchy and if this results in a unique match then all the UCTs in the BCT or TCT have been disambiguated and the singular UCTs that were generated for the BCT or TCT are removed. If there is no unique match then the BCT or TCT is discarded and further disambiguation is performed only on the UCTs.

[13] `http://developer.yahoo.com/geo/geoplanet/`

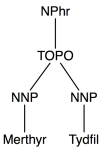

Fig. 8 The toponym “Merthyr Tydfil” with the syntactical structure and the TOPO element that is used to group NNPs that are part of a toponym.

While the initial disambiguation used simple logic to disambiguate, the remaining disambiguation methods only provide a ranking as to how likely it is that each of the UCT’s possible geocodings is the correct geocoding. The individual method’s rankings are combined and for each UCT the geocoding with the highest ranking is chosen as the final, disambiguated geocoding. The methods used by the Toponym Ontology are: shared hierarchy with previously unambiguously grounded toponyms; spatial correlation with both ambiguous candidate geocodings and disambiguated geocodings; shared feature type between the UCT and the candidate geocoding; web popularity ranking (developed by Xin Fan in University of Sheffield) and population size.

After the geocoding process the POS tagged caption is additionally tagged with the identified toponyms and their locations and is now ready to be interpreted.

5.2 Natural language parsing

This is the first of the core interpretation components and its task is to transform the POS and geotagged caption into a more generic form that can be used by the qualitative modelling component.

5.3 Syntactic toponym integration

Before the generalisation rules can be applied the toponym information needs to be integrated into the syntactical POS tagging. For each toponym the *NNP*s (proper nouns) that are identified as being part of the toponym are removed from the caption and added as children to a *TOPO* element. The *TOPO* element itself is added as the only child to a *NPhr* (noun phrase) element (fig. 8). This structure cleanly models the internal toponym structure as consisting of one or more *NNP*s and as a whole being treated as a *NPhr*. The *NPhr* is then inserted into the original caption in the location where the *NNP*s were removed.

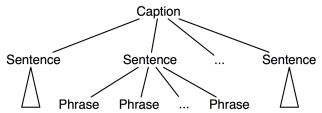

Fig. 9 The structure of the top of the qualitative model generated for a caption. The caption is split into one or more sentences and each sentence in turn is split into one or more phrases.

```
1  def generalise(sentence):
2    for rule_block in rules_list:
3      for idx in range(len(sentence)):
4        for rule in rule_block:
5          if rule.applies(sentence, idx):
6            sentence.replace(idx, rule.body,
7                             rule.head)
8            return generalise(sentence)
9    return sentence
10
11 class Rule:
12   def applies(self, sentence, start):
13     for idx in range(len(self.body)):
14       if start + idx < len(sentence):
15         if self.body[idx].pos ==
16            sentence[idx + start].pos:
17           if self.body[idx].word and
18              self.body[idx].word !=
19              sentence[idx + start].word:
20             return False
21         else:
22           return False
23       else:
24         return False
25     return True
```

Listing 1 The recursive syntactic generalisation algorithm

Head	Body
IN	(RB ”north”) (IN ”of”)
IN	(RB ”south”) (IN ”of”)
IN	(RB ”east”) (IN ”of”)
IN	(RB ”west”) (IN ”of”)
NPhr	N N
NPhr	N NPhr
NPhr	J N
NPhr	J NPhr
NPhr	DT N
NPhr	DT NPhr
NPhr	N POS N
NPhr	NPhr POS N
NPhr	N POS NPhr
NPhr	NPhr POS NPhr
NPhr	N
ConjPhr	NPhr CC NPhr
IPhr	IN NPhr
IPhr	IN CPhr

Table 6 List of generalisation rules employed by the natural language parsing to simplify the syntactical structure. Each rule consists of one or more body elements that must match part of the caption and the head is then used to replace the body elements. Where both a syntactical class and a specific word are given, both must match the part of the caption for the rule to be applied. The rules have been simplified and all types of nouns (NN, NNP, NNS, NNPS) are simply referred to as N and all types of adjectives (JJ, JJR, JJS) as J in order to keep the table simpler.

Fig. 10 Example of the generalisation structure used to transform the cardinal direction “north of” into a single preposition element. This guarantees that the cardinal directions can then be handled like any other spatial preposition.

5.4 Syntactic caption generalisation

The generalisation algorithm starts by splitting the caption along full-stops, if there are any. Each sentence is then split again at commas creating the tree structure outlined in figure 9. Each phrase as delineated by commas is then generalised using the rules shown in table 6. The generalisation algorithm (lst. 1) is implemented as a recursive function that attempts to match each rule to parts of the phrase by sliding the body elements of the rule over the phrase from right to left. If the body elements of the rule match some or all the elements of the phrase, then the matching elements are removed from the phrase and replaced by a new element defined by the head of the rule. The removed elements are added as children to the new head element and the algorithm is then called recursively with the partially generated phrase. In most cases the matching algorithm only compares the POS type, but for the cardinal direction handling the actual words are also compared (tab. 6, lines 1-4) so that they can then be treated like any other spatial preposition (fig. 10).

It is necessary that certain rules are applied before other rules are applied. For example all noun phrase generation rules need to be applied before the conjunctive phrase or prepositional phrase rules are applied. The orderings are maintained by grouping the generalisation rules and the algorithm only proceeds to processing the rules of the next group if none of the rules of the current group could be matched to the phrase. Due to the fact that apart from the partially generalised phrase no information is passed when the function is called recursively, the algorithm will always attempt to match rules from rule groups that have already been processed. While this is computationally inefficient, it simplifies the design and implementation of the algorithm and since image captions tend to be relatively short and the number of rules is relatively small, the trade-off is worthwhile.

After all rules have been applied each phrase consists of a sequence of words and root elements of generalised structures. The phrases are attached to the caption - sentence tree structure (fig. 11) and this tree structure is passed on to the qualitative modelling component.

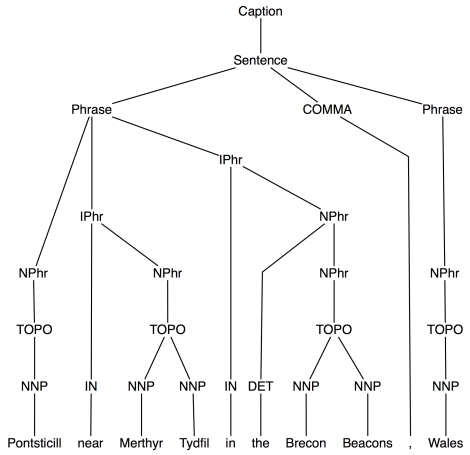

Fig. 11 The full generalisation tree structure for the caption "Pontsticill near Merthyr Tydfil in the Brecon Beacons, Wales".

5.5 Qualitative modelling

The next step is to build a qualitative spatial model from the generalised caption tree structure and this is performed in two steps. First all the spatial information is extracted from the caption tree and then the extracted qualitative model is pruned to remove those elements that do not add to the knowledge represented by the model.

The modeller starts at the top of the caption three and creates a cascading series of WEAK-AND elements for the list of sentences (fig. 12). Each WEAK-AND combines either two sentences, a sentence and another WEAK-AND or a single sentence if there is only one sentence in the caption. The WEAK-AND indicates that the spatial information of all its children is to be combined, but that the information in the left-most child is to be seen as of higher value than that of the other children. This is because in a most captions the most important spatial information will be placed in the first sentence of the caption, so that if the viewer only reads the beginning of the caption they still know where and of what the photograph was taken. For this reason the WEAK-ANDs are cascaded, representing the fact that the further away from the beginning of the caption the sentence is, the lower its relevance.

Under the WEAK-ANDs a STRONG-AND element is added for each sentence, which acts as a container for the elements generated for the individual phrases. The STRONG-AND element indicates that the information in all its children is to be treated as equally relevant. This is because each phrase was separated by commas and while the commas separate the sentence into multiple parts, these parts are all of the same conceptual level and thus to be combined as such. Finally for each phrase in a sentence a STRONG-AND element is generated (fig. 12) and added to the sentence's STRONG-AND element.

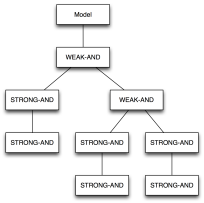

Fig. 12 The initial qualitative model generated for a caption with two sentences such as "Tree in the Brecon Beacons. Photographed near Pontsticill reservoir. I took this photo while hiking.". At this point no distinction has been made as to which phrases contain spatial information.

After this upper model is created, the qualitative modeller begins extracting the actual spatial information from the phrases. Since the information in English sentences is ordered from left to right, the modeller also extracts the information from left to right. In a first step those generalised trees that contain spatial information are extracted, which are the *NPhr*s and *IPhr*s that contain toponyms, marked by the *TOPO* element in the syntactical tree. All other information in the sentence is discarded and while in some cases this information might help to locate the image, it exceeded the scope of this work[14].

These spatial structures are integrated into the model from left to right under the phrase's STRONG-AND element. Individual noun phrases are added to the model as a toponym-element. Conjunctive phrases are added as a STRONG-AND element with the two toponyms as child elements. For spatial prepositions an element of the type of the spatial preposition is added (fig. 15).

When a prepositional element is added the `QualitativeModeller` checks whether previous qualitative model elements have been created for this phrase. If there are any such elements, then these are replaced by the new prepositional element. The replaced elements are added as the *figure* child of the preposition element (fig. 13). This works due to the left-to-right ordering of the spatial information in the caption. The reasoner assumes that the spatial preposition relates the spatial information that has already been extracted to the spatial preposition's *ground* toponym. While this heuristic fails if the toponym identification did not identify a toponym, the toponym identification is relatively robust and thus this heuristic works very well.

If the noun phrase contained in a prepositional phrase is a conjunctive phrase ("near Abercynon and Pontypridd") then the qualitative model has to be rearranged as the quan-

[14] An example of how this additional information might help is in the caption "Flowers near Stackpole Head", where Stackpole Head is a coastal headland, and the inclusion of the knowledge that the photo is of "Flowers" could restrict the probable area where the photo was taken to the land-side, whereas if the subject were "Sailing boat" then the sea-facing "near" area would be more likely. This kind of knowledge processing was decided to be beyond scope of this paper and thus the information is discarded

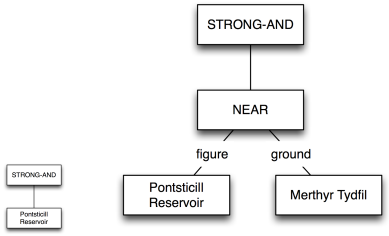

Fig. 13 Modelling of the caption "Pontsticill Reservoir near Merthyr Tydfil". The diagram on the left shows the model after "Pontsticill Reservoir" has been extracted, the diagram on the right the model after adding "near Merthyr Tydfil" and rearranging the model.

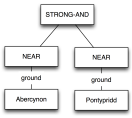

Fig. 14 The model structure generated for the phrase "near Abercynon and Pontypridd".

Fig. 15 The full qualitative model generated for a caption with two sentences such as "Tree in the Brecon Beacons. Photographed near Pontsticill reservoir".

tification only allows for a single *ground* toponym when quantifying the spatial prepositions. In order to support conjunctive phrases in prepositional phrases a separate prepositional element is created for each of the toponyms in the conjunctive phrase and the prepositional elements are combined under a STRONG-AND element, thus the phrase "near Abercynon and Pontypridd" is modelled as "near Abercynon and near Pontypridd" (fig. 14). While it might be argued that these are not completely equivalent, they are sufficiently similar to not distort the result too strongly.

As the `QualitativeModeller` traverses the caption from left to right a potentially very complex model (fig. 15) is built that very often contains elements that add no further information to the model as a whole. In the next step a number of heuristics are used to prune the model.

5.5.1 Pruning

As stated the aim of the pruning is to reduce the amount of processing required in the further stages. The pruner works bottom-up from left-to-right and at each node it uses a set

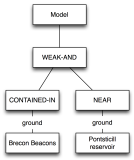

Fig. 16 The pruned qualitative model for the caption "Tree in the Brecon Beacons. Photographed near Pontsticill reservoir".

of rules to determine whether the node is required looking both at the node-type and its children. The simplest rule removes those STRONG-AND and WEAK-AND elements that have only one child. The child is moved up the hierarchy and takes the STRONG-AND or WEAK-AND's place. If a preposition element has children in both the *figure* and *ground* links then the information in the *ground* links is discarded along with the preposition element and the *figure* child takes the place of the preposition element. The reason for this is that due to the basic nature of caption language the *figure* elements will always be more precise than the *ground* elements. Thus if both are available the spatial information provided by the *ground* elements and the spatial preposition only add noise to the final model and will not improve the quality of the model (fig. 16) and consequently the `ground` element is pruned.

5.6 Quantification

The pruned qualitative model created in the previous step now has to be augmented with quantitative information to transform it into a vague-field which can then be translated into the image's footprint. The quantitative data is drawn from the experiment described in section 3.2, which is stored and manipulated using the vague-field model (sect. 4). The quantification process transforms the qualitative model into a single unified field that describes the likely area where the photograph was taken.

The quantification algorithm (lst. 2) recursively traverses the qualitative model tree top-down, left-to-right, first descending to the leaf elements and then adding and integrating the vague fields as it moves back up the qualitative model tree. Depending on the element type and any children the element may have, vague-fields are either created or combined.

5.6.1 Toponym model elements

For toponym model elements the algorithm simply instantiates a toponym field for the toponym associated with the model element. A toponym field is a circular, crisp vague-field used to approximate the toponyms shape. No further processing is performed, but it is necessary to create the to-

```
def quantify(node):
  for child in node.children():
    quantify(child)
  if node.type == WEAK_AND:
    node.field = node.children().first.
        field.combine(node.children(). rest.fields,
                      0.5)
  elif node.type == STRONG_AND:
    node.field = node.children().first.
        field.combine(node.children(). rest.fields)
  elif node.type == PREPOSITION:
    node.field = SparsePointMeasurementField(
                   node.ground().toponym())
  elif node.type == TOPONYM:
    node.field = CrispField(node.toponym())
```

Listing 2 The recursive quantification algorithm first descends top-down into the models child nodes and then constructs the field structures bottom-up by either instantiating the respective field or invoking the combine operation. The `children` function returns all child nodes, while the other functions (`ground`, `start`, `end`, `toponym` return specific child nodes. The code shown here does not include the caching functionality which is described later.

ponym field to allow the algorithm to deal with all elements without having to differentiate between toponyms and fields.

5.6.2 Preposition model elements

The preposition model elements that are not removed by the model pruning are those that have no child element linked via the *figure* relation. The *ground* relation will link to a toponym element that due to the bottom-up nature of the algorithm will have been transformed into a toponym field. From the toponym field only the toponym's coordinates are used to instantiate the prepositional field, all other data from the toponym field is discarded for two reasons. First the spatial preposition application data was acquired based on point-like toponyms (sect. 3.2) and no data was acquired that would make it possible to model how the field should be instantiated if the toponym is of extended shape. Second the toponym fields are arbitrary approximations as the geo-data represents them as points, even though they are polygonal in shape, thus using them to distort the field would only introduce error into the field and not improve the quality of the result.

In the instantiation of the prepositional vague-field model the reasoner will first check if there is a cached version of the necessary vague-field available as this reduces the processing time significantly. If there is then the field is instantiated from cache and anchored using the toponym's coordinates. However if there is no cached version available then the field is instantiated using the sparse-measurements method (sect. 8.2) and after it has been instantiated a cached version is created and stored both on disk to speed up future processing.

5.6.3 Combinatorial model elements

Combinatorial model elements (STRONG-AND and WEAK-AND) are handled by applying the `scalar` and `combine` operations to the fields created by their child elements (lst. 2). In the case of the STRONG-AND element the fields are simply merged using the first field as the base field and then using the `combine` method to add the other fields. If the element is a WEAK-AND then the first field is left as it is, but for all remaining fields the `scalar` operation is applied, with each field being multiplied by a factor of 0.5 to indicate that while the information contained in them is to be considered, it is not as important and highly valued as the information in the first field, then the fields are merged using the `combine` operation.

After the algorithm has integrated all the individual fields and processed each of the qualitative model's nodes the result is a single continuous field where each field cell represents no longer the applicability of a given spatial preposition, but the likelihood that the photograph was taken at that location or is of an object that is at that location. At this point the translation from the qualitative, spatio-linguistic representation in the image caption to the quantitative, computational model is complete. The next step of creating a crisp footprint is simply an arbitrary decision by the algorithm as to which parts of the field are likely to be areas where the photograph was taken and which are not.

5.7 Footprint calculation

The final processing step is the calculation of the crisp footprint. The footprint calculation applies the `crisp` operation (sect. 8.5) to the final combined field that was created in the previous step. After the active contour has calculated the footprint polygon this must be projected to the desired output coordinate system. The data can be projected back into any coordinate system specified when invoking the spatial language interpreter and if none is specified then by default the coordinates will be projected back into WGS84 coordinates. The re-projection introduces some errors and skews the resulting polygon slightly, but as integration into other GI systems such as web-mapping clients requires WGS84 coordinates this is unavoidable.

6 Evaluation experiment

The difficulty with evaluating an interpretation such as the image's calculated footprint is how to determine a baseline against which the interpretation can be compared. Using an existing set of geo-referenced captions (such as the Geograph data-set) would allow testing of how often the image

coordinates lie within the calculated footprint. Two problems with this approach are the lack of negative evidence and the difficulty with interpreting the result. The lack of negative evidence means that it is hard to determine whether the footprint is too large, because none of the captions use a phrase such as "not near Brecon". Thus the perfect footprint would be all-encompassing as then the image coordinates would always be inside the footprint. The second problem is linked to this one and is that if the footprint is not all-encompassing what percentage of captions is it acceptable to have outside the footprint. How many of the captions could be considered borderline or wrong in themselves and how often should one assume that the geocoding is incorrect. Effectively such an approach would only replace one difficult problem with a different equally difficult one.

As an alternative, using a much smaller set of captions, it is possible to use human annotators to create a "gold-standard" set of footprints. The reasoner's footprints can then be compared to the annotators' footprints using qualitative comparisons, and also using another set of evaluators who compare the footprints and rate how well the footprints fit the captions. This experimental approach creates a baseline of how highly the human evaluators rate the human annotators' footprints, against which the evaluators' ratings of the computer-generated footprints can be compared.

The results of these evaluations highlighted some of the problems with the behaviour of the language interpretation which led to some modifications (sect. 7) that were made in order to create results that are closer to those created by humans. In this section these shortcomings will be highlighted and the next section will illustrate the modifications made to the algorithms.

6.1 Baseline creation

The baseline human-generated footprints were created by three annotators. They were shown eight image captions plus maps of the areas where the photograph was taken and on the maps the toponyms mentioned in the caption plus toponyms in the surrounding area. The annotators were then instructed to mark out the area on the maps where they thought it was likely that the photograph had been taken based on the caption. A total of 24 footprints (8 per annotator) were created like this using the spatial prepositions "near", "north of", "east of ", and "between"[15]. For each spatial preposition two captions plus maps were shown and the annotators had no knowledge of the areas where the photographs were taken. The baseline outlines were then digitised by scanning the hand-drawn maps and then tracing the outlines of the marked areas and then the outlines were filled in.

[15] The results for "east of" will not be reported as they are analogous to the "north of" results and the "between" results due to space constraints.

	Annotator 1	Annotator 2	Annotator 3	Algorithm
Caption 1	6 / 3	5 / 5	7 / 2	2 / 2
Caption 2	7 / 2	6 / 2	7 / 2	2 / 2

Table 7 Median and inter-quartile ranges for the two captions "Pond near High Buston" (caption 1) and "Hopwell farm near Castle Headingham, Essex" (caption 2).

6.2 Experimental design

A total of 85 participants were recruited from undergraduates and staff at Cardiff University. They were shown a set of 8 questions and for each question the image caption plus the three annotator-created footprints and the computer-generated footprint were shown. The footprints were superimposed on the same maps that were shown to the annotators. Below each map a 9-point Likert-type scale was provided and the participants were instructed to use these to rate how well the footprints matched the caption, with a rating of 1 as no match and a rating of 9 a perfect match. As with the earlier experiment the evaluation is based on median values and inter-participant agreement. As in the earlier experiment an inter-participant agreement is based on the inter-quartile range[16]. Additionally the reasoner's footprint is classified into three classes ("achieve", "almost achieve", "fail") for each participant as to whether the footprint achieves the goal of being as good as the footprints produced by the annotators.

6.3 Near

The two captions used in the questionnaire were "Pond near High Buston" and "Hopwell farm near Castle Headingham, Essex". Table 7 shows median values and inter-quartile range for the three human annotators and the reasoner's results. As is clearly visible the reasoner's ratings are much lower than all of the human annotators for both captions (tab. 9). At the same time the median and inter-quartile ranges for the human annotators also show that people do not agree with each other in their interpretations (inter-quartile ranges of 2 or higher). If the median values are reduced to a three-level scale then only half the human annotators shapes have a high median rating (tab. 8), clearly illustrating the problem of creating an answer that is acceptable to a large group of people, but it also illustrates that the reasoner does not perform well at all.

Looking at the shapes created by the human annotators and the reasoner (fig. 17) it is clear that there are further constraints that the annotators take into account that the reasoner does not know of. The extent of the reasoner's shape

[16] 0 and 1 - high agreement, 2 - medium agreement, 3 or higher - low agreement

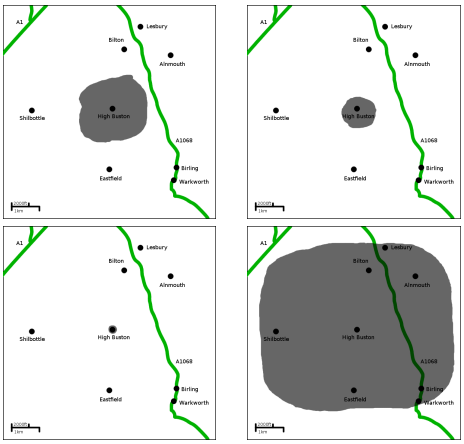

Fig. 17 The four maps used for the evaluation of "near" in the caption "Pond near High Buston". The bottom-right images is the computer-generated outline, while the other three have been produced by the human annotators (Annotator 1 top-left, Annotator 2 top-right, annotator 3 bottom-left).

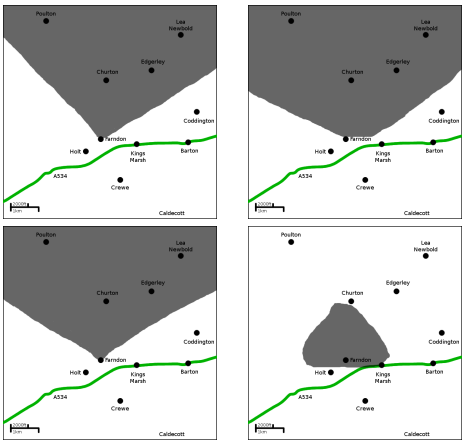

Fig. 18 The four maps used for the evaluation of "north of" in the caption "North of Farndon". The bottom-right image is the computer-generated outline, while the other three have been produced by the human annotators (Annotator 1 top-left, Annotator 2 top-right, annotator 3 bottom-left).

	Annotator 1	Annotator 2	Annotator 3	Algorithm
Caption 1	2 / 1	2 / 2	3 / 1	1 / 1
Caption 2	3 / 1	2 / 1	3 / 1	1 / 1

Table 8 Median and inter-quartile ranges for the two captions "Pond near High Buston" (caption 1) and "Hopwell farm near Castle Headingham, Essex" (caption 2) reduced to a 3-level scale.

	Annotator 1	Annotator 2	Annotator 3	Algorithm
Caption 1	6 / 2	6 / 2	5 / 3	4 / 3
Caption 2	8 / 1	6 / 3	6 / 2	4 / 3

Table 10 Median and inter-quartile ranges for the two captions "North of Farndon" (caption 1) and "Pound farmhouse just north of Rayne, Essex" (caption 2).

	Achieve	Almost	Fail
Caption 1	0.12	0.06	0.82
Caption 2	0.1	0.13	0.84

Table 9 Percentages of evaluators' answers that fall into the categories "as good as human", "almost as good as human", and "not as good" for the two captions "Pond near High Buston" (caption 1) and "Hopwell farm near Castle Headingham, Essex" (caption 2).

	Achieve	Almost	Fail
Caption 1	0.33	0.05	0.63
Caption 2	0.18	0.14	0.67

Table 11 Percentages of evaluators' answers that fall into the categories "as good as human", "almost as good as human", and "not as good" for the two captions "North of Farndon" (caption 1) and "Pound farmhouse just north of Rayne, Essex" (caption 2).

matches the distances seen in the Geograph analysis (Hall and Jones (2008)) with a distance of not quite three kilometres from the centre to the edge of the area, but in the given examples the existence of places closer to the centre act as a constraint as points around these places would be referred to as being "near" those places instead. The human annotators' shapes differ from each other as well, so clearly there are further possibly more personal constraints and interpretations which are harder to quantify, but it is this constraint imposed by the proximal places that needs to be considered to improve the reasoner's ratings.

North of

The two captions used in the questionnaire were "North of Farndon" and "Pound farmhouse just north of Rayne, Essex". As the median values (tab. 10) show, the reasoner's results are better than for "near", but still not as good as the annotators' results (tab. 11). The reasoner's ratings are better for the first caption "North of Farndon" and the lower ratings for the second caption are probably due to the fact that the second caption specifies the area as "just north" and the reasoner does not know about such modifiers. An interesting effect is that the ratings for the annotators are higher for the second caption. It seems that with a more strongly constrained caption ("just north" instead of "north") the reduced vagueness brings people's mental models closer together improving their ratings. Nevertheless as in the "near" case the inter-participant agreement is not high with inter-quartile ranges of 2 or higher for all results.

One aspect that is instantly visible is that the reasoner's "north" shape is strange (fig. 18) and that it overlaps into the area that most people would refer to as "south". The reasons for this are that a distance-factor is included to model the distance-effect seen in the analysis of the Geograph data (Hall and Jones (2008)) but not in the human-subject exper-

iment upon which the field is based. The decision was made to include this effect in the interpretation, but it does lead to the slightly strange shape which is not seen as correct by the evaluators. Especially in the first caption when comparing the annotators' shapes with the algorithmic shape then it is clear that distance is not seen as an important factor and the field generated should not include a distance factor. On the other hand the "just" modifier in the second caption provides a very strict distance constraint which should if possible also be considered.

6.4 Conclusion

The primary result of the evaluation is that some of the assumptions that were made when modelling the spatial prepositions' quantitative aspects do not match with what people produce and expect. Due to these modelling errors the results are still a long way from achieving our goal of being "as good as humans", but the next section will present some modifications that should take us closer to our goal.

7 Modifications

The algorithms and data-acquisition experiments were designed based on the initial data-mining experiments that provided a general overview of how the quantitative aspects of spatial language worked. The one aspect that was not clearly discernible from the data-mining results was how much of the quantitative effects were a-priori effects inherent to the spatial language and how much were a-posteriori effects caused by contextual aspects such as the decision process leading to one caption being chosen from the range of possible descriptions. The data-acquisition experiments and the algorithms were designed assuming an a-priori view, but as the evaluation experiment shows some of the effects are a-posteriori or at least influenced by further constraints. Taking the results from the evaluation experiment the source data and algorithms have been modified to take this new knowledge into account and thus improve the quality of the generated footprints and also to illustrate that the reasoner can easily be updated when new knowledge becomes available. The modified footprints have not been re-evaluated using the annotator base-line, instead the new footprints are qualitatively compared to the annotators' footprints.

7.1 Crisping

To enable the modifications the crisping algorithm had to be partially re-designed (eq. 5) and now uses the scalar values directly without translating into the vector representation. This change also meant that the weights attached to each energy component have changed and the values used in this section are $\alpha = 0.0001$, $\beta = 1$ and $\gamma = 1$.

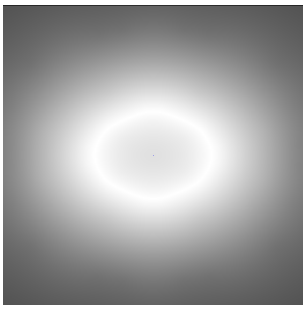

Fig. 19 The field that has been transformed using equation 7 with $\delta = 0.8$. The lighter the image, the lower the field energy at that point. Clearly visible as a white ring is the trough formed by the transformation.

$$E_{snake} = \alpha \cdot E_{int} + \beta \cdot E_{field} + \gamma \cdot E_{constraint} \tag{5}$$

The updated internal energy E_{int} is calculated according to equation 6, where *angle* is the angle at point p_i in the triangle defined by p_{i-1}, p_i, p_{i+1} and *dist* the difference between the distances from p_i to p_{i-1} and from p_i to p_{i+1} (fig. 28 shows the relative locations of the three control points). If the *angle* is greater than 60° or the *dist* is less than 1 then a hard limit is enforced to avoid the snake becoming overly angular and the control points merging.

$$\begin{aligned} E_{int} &= E_{angle} + E_{dist} \\ E_{angle} &= \begin{cases} |180 - angle| & \text{if } |180 - angle| \leq 60 \\ 1000000 & \text{if } |180 - angle| > 60 \end{cases} \\ E_{dist} &= \begin{cases} |dist| & \text{if } |dist| > 1 \\ 1000000 & \text{if } |dist| \leq 1 \end{cases} \end{aligned} \tag{6}$$

The contraction energy is replaced by an α-cut like preprocessing step where the values of the main field are transformed as described in equation 7 (fig. 19). The δ value replaces the α, but unlike the crisp outline generated by the α-cut the result is an energy-trough into which the active contour will roughly settle. Once the active contour has settled into that trough the constraint and internal energies deform it and produce the final, smooth footprint.

$$E_{field} = \left| \delta - E_{field} \right| \tag{7}$$

Finally the constraint energy $E_{constraint}$ is used to improve the quality of the "near" fields, by allowing competing interpretations as will be explained in the next section.

7.2 Near

As was clearly illustrated in the previous section the basic problem with the interpretation of the spatial preposition

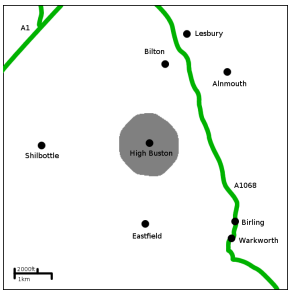

Fig. 20 The footprint for "near" produced by the updated field definition and crisping algorithm.

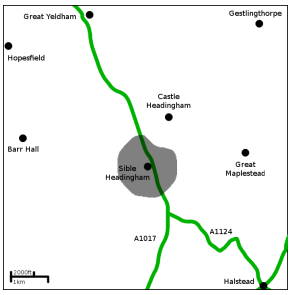

Fig. 21 The footprint for "near" demonstrates the constraint influence exerted by near field around "Castle Headingham", which creates a dent in the active contours outline.

"near" is that it does not take into account the negative evidence provided by the photo location being "near" another place. The updated crisping algorithm provides a constraint energy component that allows for this influence, which is defined as the sum of all constraint fields (eq. 8). There is no normalisation as that could diminish the negative evidence at one location solely because at another location two bits of negative evidence overlapped. A high concentration of negative evidence at one location does not make the negative evidence at another location any less valid, which is what normalisation would result in.

$$E_{constraint} = \sum_{i=0}^{n} \text{ConstraintField}_i \tag{8}$$

Initially the list of constraint fields is empty, however when the quantifier processes a "near" prepositional element, in addition to creating the "near" field for that element it also adds to the constraint energy list. It does this by querying a reverse geocoder developed in the Tripod project to retrieve a set of towns and villages around the point where the field is initialised. For each of the surrounding places a "near" field is created and added to the list of constraint energies. Thus when creating the main field the quantifier is effectively also creating a list of fields that contain alternative possible descriptions.

The updated crisping algorithm has been used both with the constraint energy (fig. 21) and without (fig. 20) and as the figures show the footprints are much closer to the annotator-created footprints, but further testing is required to determine which of the two footprints would be rated higher.

7.3 Cardinal directions

Modifying the cardinal directions support does not require any changes to the algorithms, only the field definitions need to be updated to remove the distance-weighting from the field definitions (fig. 22). Additionally a rule to filter the field so that the footprint does not overshoot into the opposite half-plane was tested, but as the overshoot is very small and to account for errors both in the original captioning process and in the geocoding the filter was not kept in the end.

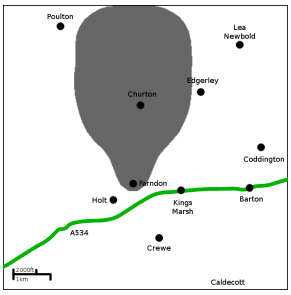

Fig. 22 The footprint for "north" produced by the updated field definition.

As figure 22 shows the new footprint is much closer to the footprints created by the annotators which again should lead to a corresponding improvement in the ratings. The new footprint is also vaguer than the original footprints as they cover a larger area and the reasoner is thus not as definite about where it believes the photograph was taken, but that seems to be the preferred shape by both annotators and evaluators.

8 Conclusion

In this paper we presented a system that makes it possible to determine images' locations based on the spatial information contained in their captions. This enables images that do not have explicit locational meta-data to be included in Geographic Information Retrieval applications.

Our approach uses a field-based model to represent the vagueness that is implicit in spatial language. The quantitative data that underpins these fields is based on data acquired using human-subject experiments. These are combined with a spatio-linguistic reasoner to enable the translation of captions such as "Bridge in Cambridge" to a polygonal footprint representing the likely area where the photograph was taken. This approach has been evaluated by comparing the generated footprints to footprints created by human annotators and this evaluation highlighted a few shortcomings which we have addressed to bring us closer to our goal of

creating footprints that are as good as those created by human annotators.

Future work will focus on making the spatio-linguistic reasoner more powerful and resilient so that it can successfully interpret a wider range of image captions, investigating the issues of conflicting descriptions identified in the evaluation of the "near" evaluation data, and determining quantitative models for further spatial prepositions. On the evaluation side the updated footprints need to be re-evaluated against the human annotators footprints to determine whether the modifications made actually improve the results.

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.

We would also like to thank the two reviewers, whose comments and suggestions helped focus the ideas presented in this paper.

References

Ahlqvist O, Keukelaar J, Oukbir K (1998) Using rough classification to represent uncertainty in spatial data. In: Proceedings of the SIRC Colloquium, pp 1–9

Altman D (1994) Fuzzy set theoretic approaches for handling imprecision in spatial analysis. International Journal of Geographical Information Science 8(3):271–289

Andogah G, Bouma G, Nerbonne J, Koster E (2008) Placename ambiguity resolution. In: LREC Workshop on Methodologies and Resources for Processing Spatial Language

Bennet B (2001) Application of supervaluation semantics to vaguely defined spatial concepts. In: Spatial Information Theory. Foundations of Geographic Information Science : International Conference, COSIT 2001 Morro Bay, CA, USA, September 19-23, 2001. Proceedings, pp 108–123

Bennett B, Agarwal P (2007) Semantic categories underlying the meaning of 'place'. In: Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, pp 78–95

Bittner T, Stell J (2003) Stratified rough sets and vagueness. In: Spatial Information Theory, Springer Berlin / Heidelberg, pp 270–286

Bowerman M, Choi S (2003) Space under construction: Language-specific spatial categorization in first language acquisition. In: Gentner D, Goldin-Meadow S (eds) Language in mind, MIT Press, pp 387–428

Brown P (1994) The ins and ons of tzeltal locative expressions. Linguistics 32:743–790

Burghardt D (2005) Controlled line smoothing by snakes. GeoInformatica 9(3):237–252

Buscaldi D, Rosso P (2008) Map-based vs. knowledge-based toponym disambiguation. In: Proceeding of the 2nd international Workshop on Geographic information Retrieval.GIR'08, pp 19–22

Chomsky N (1965) Aspects of the Theory of Syntax. MIT press

Clementini E, Felice PD (1996) An algebraic model for spatial objects with indeterminate boundaries. In: Geographic Objects with Indeterminate Boundaries, Taylor and Francis, pp 155–169

Clementini E, Felice PD (1997) Approximate topological relations. International Journal of Approximate Reasoning 16(2):173–204

Cohn A, Gotts N (1996a) The 'egg-yolk' representation of regions with indeterminate boundaries. In: Proceedings, GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries, Francis Taylor, pp 171–187

Cohn A, Gotts N (1996b) Representing spatial vagueness: A merological approach. In: KR'96: Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, pp 230–241

Couclelis H (1992) People manipulate objects (but cultivate fields): Beyond the raster-vector debate in gis. In: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, vol 639/1992, Springer Berlin / Heidelberg, pp 65–77

Couclelis H, Gottsegen J (1997) What maps mean to people: Denotation, connotation, and geographic visualization in land-use debates. In: Spatial Information Theory A Theoretical Basis for GIS (COSIT'97), vol 1329/1997, Springer Berlin / Heidelberg, pp 151–162

Coventry K, Prat-Sala M, Richards L (2001) The interplay between geometry and function in the comprehension of over, under, above and below. Journal of Memory and Language 44(3):376–398

Cunningham H, Maynard D, Bontcheva K, Tablan V (2002) 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, pp 168–175

Edwardes A, Purves R (2007) A theoretical grounding for semantic descriptions of place. Lecture Notes in Computer Science 4857:106

Egenhofer M (1991) Reasoning about binary topological relations. In: Second Symposium on Large Spatial Databases, Lecture Notes in Computer Science, vol 525, Springer-Verlag, pp 143–160

Erwig M, Schneider M (1997) Partition and conquer. In: COSIT '97: Proceedings of the International Conference on Spatial Information Theory, Springer-Verlag London,

pp 389–407

Fabrikant S, Buttenfield B (2001) Formalizing semantic spaces for information access. Annals of the Association of American Geographers 91(2):263–280

Fisher P (2000) Sorites paradox and vague geographies. Fuzzy Sets and Systems 113(1):7–18

Fisher P, Wood J, Cheng T (2004) Where is helvellyn? fuzziness of multi-scale landscape morphometry. Transactions of the Institute of British Geographers 29(1):106–128

Fisher PF, Orf TM (1991) An investigation of the meaning of near and close on a university campus. Computers, Environment and Urban Systems 15(1-2):23–35, DOI DOI:10.1016/0198-9715(91)90043-D

Frank AU, Raubal M (1999) Formal specification of image schemata – a step towards interoperability in geographic information systems. Spatial Cognition and Computation 1(1):67–101

Friedman C, Kra P, Yu H, Krauthammer M, Rhzetsky A (2001) Genies: a natural-language processing system for the extraction of molecular pathways from journal articles. Bioinformatics 17(1):74–82

Fuhr T, Socher G, Scheering C, Sagerer G (1995) A three-dimensional spatial model for the interpretation of image data. In: IJCAI-95 Workshop on the Representation and Processing of Spatial Expressions, pp 93–102

Gahegan M (1995) Proximity operators for qualitative spatial reasoning. In: Spatial Information Theory A Theoretical Basis for GIS, Springer Berlin / Heidelberg, pp 31–44

Gapp K (1994) Basic meanings of spatial relations: Computation and evaluation in 3d space. In: National Conference on Artificial Intelligence, pp 1393–1398

Gärdenfors P (2000) Conceptual spaces: The geometry of thought. MIT Press

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

Goodchild M (1992) Geographical data modeling. Computers & Geosciences 18(4):401–408

Goodchild M, Hill L (2008) Introduction to digital gazetteer research. International Journal of Geographic Information Science 22(10):1039–1044

Guo Q, Liu Y, Wieczorek J (2008) Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. International Journal of Geographical Information Science 22(10):1067–1090

Güting R, Schneider M (1993) Realms: A foundation for spatial data types in database systems. In: Proceedings of the 3rd International Symposium on Large Databases, pp 33–44

Hall M, Jones C (2008) Quantifying spatial prepositions: an experimental study. In: Proceedings of the ACM GIS'08, pp 451–454

Hall MM, Jones CB (2009) Initialising and terminating active contours for vague field crisping. In: GISRUK 2009, pp 395–397

Hall S (1980) Encoding/decoding. In: for Contemporary Cultural Studies C (ed) Culture, Media, Language: Working Papers in Cultural Studies 1972-79, London: Hutchinson, pp 128–138

Hengl T (2007) A practical guide to geostatistical mapping of environmental variables

Herskovits A (1986) Language and Spatial Cognition: An Interdisciplinary Study of Prepositions in English. Cambridge University Press

Horvath P, Jermyn I, Kato Z, Zerubia J (2009) A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction. Pattern Recognition 42(5):699–709

Hwang S, Thill JC (2005) Modeling localities with fuzzy sets and gis. Fuzzy Modeling with Spatial Information for Geographic Problems pp 71–104

Johnson M (1987) The Body in the Mind. University of Chicago Press

Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. International Journal of Computer Vision 1(4):321–331

Kemmerer D (2006) The semantics of space: Integrating linguistic typology and cognitive neuroscience. Neuropsychologia 44(9):1607–1621

Kemmerer D, Tranel D (2000) A double dissociation between linguistic and perceptual representations of spatial relationships. Cognitive Neuropsychology 17(5):393–414

Klippel A, Montello D (2007) Linguistic and nonlinguistic turn direction concepts. In: Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, pp 354–372

Klir G, Yuan B (1995) Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice-Hall

Krige D (1951) A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Chemical, Metallurgical and Mining Societe 52(119-139)

Kuhn W (2002) Modeling the semantics of geographic categories through conceptual integration. In: Geographic Information Science: Second International Conference, GIScience 2002, pp 108–118

Kulik L (2001) A geometric theory of vague boundaries based on supervaluation. In: Spatial Information Theory. Foundations of Geographic Information Science : International Conference, COSIT 2001, Springer Berlin / Heidelberg, pp 44–59

Lakoff G, Johnson M (1980) Metaphors We Live By. The University of Chicago Press

Lam KM, Yan H (1994) Fast greedy algorithm for active contours. Electronics Letters 30(1):21–23

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

Laurini R, Pariente D (1996) Towards a field-oriented language: First specifications. In: Geographic Objects with Indeterminate Boundaries, Taylor and Francis, pp 225–236

Leidner J (2007) Toponym resolution in text: Annotation, evaluation and applications of spatial grounding of place names. PhD thesis, School of Informatics, Edinburgh, UK

Levinson S (2003) Space in language and cognition: Explorations in cognitive diversity. Cambridge: CUP

Levinson S, Kita S, Haun D, Rasch B (2002) Returning the tables: language affects spatial reasoning. Cognition 84(2):155–188

Li P, Gleitman L (2002) Turning the tables: language and spatial reasoning. Cognition 83(3):265–294

Liu Y, Goodchild M, Guo Q, Tian Y, Wu L (2008) Towards a general field model and its order in gis. International Journal of Geographical Information Science 22(6):623–643

Liu Y, Guo Q, Wieczorek J, Goodchild M (2009) Positioning localities based on spatial assertions. International Journal of Geographical Information Science 23(11):1471–1501

Liu Y, Yuan Y, Xiao D, Zhang Y, Hu J (2010) A point-set-based approximation for areal objects: A case study of representing localities. Computers, Environment and Urban Systems 34(1):28–39

Lodge D (1984) Small World. Penguin Books

Mark D (1989) Cognitive image-schemata for geographic information: Relations to user views and gis interfaces. In: Proceedings GIS/LIS’89, pp 551–560

Mark D, Frank A (1995) Experiential and formal models of geographic space. Environment and Planning 23:3–24

Mark D, Turk A, Stea D (2007) Progress on yindjibarndi ethnophysiography. In: Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings, pp 1–19

Matheron G (1962) Traité de géostatistique appliquée. Mémoires du Bureau de Recherches Géologiques et Minières 14

Miller G, Johnson-Laird P (1976) Language and Perception. Cambridge University Press

Morrow D, Clark H (1988) Interpreting words in spatial descriptions. Language and Cognitive Processes 3:275–291

Mukerjee A, Gupta K, Nautiyal S, Singh M, Mishra N (2000) Conceptual description of visual scenes from linguistic models. Image and Vision Computing 18(2):173–187

Parsons S (1996) Current approaches to handling imperfect information in data and knowledge bases. Knowledge and Data Engineering 3(8):353–372

Pfoser D, Tryfona N, Jensen C (2005) Indeterminacy and spatiotemporal data: Basic definitions and case study. GeoInformatica 9(3):211–236

Power C, Simms A, White R (2001) Hierarchical fuzzy pattern matching for the regional comparison of land use maps. International Journal of Geographical Science 15:77–100

Purves R, Clough P, Joho H (2005) Identifying imprecise regions for geographic information retrieval using the web. In: GISRUK’05, pp 313–318

Randell D, Cui Z, Cohn A (1992) A spatial logic based on regions and connection. In: KR’92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Converence, Morgan Kaufmann, pp 165–176

Raubal M, Worboys M (1999) A formal model of the process of wayfinding in built environments. In: Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science: International Conference COSIT’99, Stade, Germany, August 1999. Proceedings, Springer Berlin / Heidelberg, Lecture Notes in Computer Science, vol 1661/1999, p 748

Ravin Y, Wacholder N (1996) Extracting names from natural-language text. Tech. Rep. Report 20338, IBM Research

Robinson V (2000) Individual and multipersonal fuzzy spatial relations acquired using human–machine interaction. Fuzzy Sets and Systems 113(1):133–145

Robinson V (2003) A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. Transactions in GIS 7(1):3–30

Sapir E (1929) The status of linguistics as a science. Language 5

Schneider M (1996) Modelling spatial objects with undetermined boundaries using the realm/rose approach. In: Geographic Objects with Indeterminate Boundaries, vol 2, Taylor and Francis, pp 141–152

Schneider M (2000) Finite resolution crisp and fuzzy spatial objects. In: International Symposium on Spatial Data Handling, pp 3–17

Schneider M (2001) A design of topological predicates for complex crisp and fuzzy regions. In: Conceptual Modeling - ER 2001, Springer Berlin / Heidelberg, Lecture Notes in Computer Science, vol 2224/2001, p 103

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

Smart P, Jones C, Twaroch F (2010) Multi-source toponym data integration and mediation for a meta-gazetteer service. In: Geographic Information Science, Springer Berlin / Heidelberg, Lecture Notes in Computer Science, vol 6292, pp 234–248, DOI 10.1007/

978-3-642-15300-6_17
Smith B, Varzi A (1997) Fiat and bona fide boundaries: Towards an ontology of spatially extended objects. In: Spatial Information Theory A Theoretical Basis for GIS (COSIT'97), Lecture Notes in Computer Science, vol 1329, pp 103–119
Smith D, Crane G (2001) Disambiguating geographic names in a historical digital library. In: Research and Advanced Technology for Digital Libraries: Fifth Europea Conference (ECDL 2001), pp 127–136
Srihari R, Rapaport W (1990) Combining linguistic and pictorial information: Using captions to interpret newspaper photographs. In: Current Trends in SNePS — Semantic Network Processing System, no. 437/1990 in Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp 85–96
Steiniger S, Meier S (2004) Snakes: a technique for line smoothing and displacement in map generalisation. In: ICA workshop on generalisation and multiple representation
Talmy L (1983) How language structures space. In: Spatial Orientation, 225-282, New York: Plenum, pp 225–282
Tang X (2004) Spatial object model[l]ing in fuzzy topological spaces : with applications to land cover change. PhD thesis, University of Twente, Enschede
Terzopoulos D (1986) Regularization of inverse visual problems involving disontinuities. IEEE Transactions PAMI-8 p 413
Tversky B, Lee P (1998) How space structures language. In: Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, p 157
Vorwerg C, Rickheit G (1998) Typicality effects in the categorization of spatial relations. In: Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, Lecture Notes in Computer Science, vol 1404/1998, Springer Berlin / Heidelberg, pp 203–222
Wang F, Hall G (1996) Fuzzy representation of geographical boundaries in gis. International Journal of Geographical Information Science 10(5):573–590
Whorf B, Carroll J, Chase S (1956) Language, thought, and reality: Selected writings of Benjamin Lee Whorf. MIT press Cambridge, MA
Wieczorek J, Guo Q, Hiimans R (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information Science 18(8):745–767
Winter S (2000) Uncertain topological relations between imprecise regions. International Journal of Geographical Information Science 14(5):411–430
Worboys M (2001) Nearness relations in environmental space. International Journal of Geographic Information Science 15(7):633–651
Worboys M, Duckham M, Kulik L (2004) Commonsense notions of proximity and direction in environmental space. Spatial Cognition & Computation 4(4):285–312
Xie X, Mirmehdi M (2006) Magnetostatic field for the active contour model: A study in convergence. In: Proceedings of the 17th British Machine Vision Conference, pp 127–136
Yamada A, Yamamoto T, Ikeda H, Nishida T, Doshita S (1992) Reconstructing spatial image from natural language texts. In: Proc. COLING-92, vol 4, pp 1279–1283
Zadeh L (1965) Fuzzy sets. Information and Control 8:338–353

Appendix A - Technical details of the vague-field

8.1 Definition

Conceptually the vague-field is a two-dimensional, unbounded, continuous scalar field defined on a external coordinate system. Computationally it is impossible to store an unbounded, continuous field, therefore in its internal representation the vague-field is a bounded, discretised, floating-point field which can easily be stored in a two-dimensional, floating-point matrix. This matrix which forms the foundation for the vague-field is augmented with further attributes that are required for the instantiation and processing of the vague-field.

To enable the translation between the external coordinate system and the internal matrix representation the field stores an external and an internal anchor location. The external anchor represents the location that the vague-field is defined as being relative to in the external coordinate system. For the spatial preposition fields this is the location of the *ground* toponym, such as the location of "Cardiff" in the case of the vague field for "near Cardiff". The internal anchor represents the point where the field is attached to the external anchor location. The values in the internal matrix are always to be interpreted as specifying the vague phenomenon's applicability relative to this internal anchor location. The internal and external anchors are used in the `read` function to translate between the external and internal coordinate systems (sect. 8.3).

8.2 Instantiation

The experiment described in the previous section resulted in a sparse set of measurement points and an applicability value for each measurement point. An interpolation using ordinary kriging is used to transform these point measurements into the continuous field representation. Ordinary

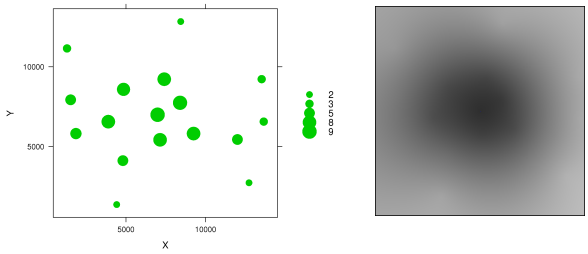

Fig. 23 The source point measurement data and the field calculated using ordinary kriging. The darker the field, the higher the applicability value at that point.

kriging was developed in the geostatistics field to estimate the distribution of natural resources based on a set of point measurements (Krige (1951); Matheron (1962); Hengl (2007)). To calculate the vague field a grid is placed over the area defined by the measurement points and the extent of each grid cell defined by the field's desired scale. The interpolated value for each cell is then calculated using a weighted average as shown in equation 9. The advantage of ordinary kriging over other distance-based interpolations is that the weighting values λ are automatically derived from the values and spatial distribution of the measurement points p (fig. 23). The interpolated results are in the range $[1,9]$ and in a final step are normalised to the $[0,1]$ range (eq. 10) where $kriging[x,y]$ is the result matrix produced by the kriging algorithm.

$$kriging[x,y] = \sum_{i=0}^{n} \lambda \cdot \text{value}\,(p_i) \tag{9}$$

$$values[x,y] = \frac{kriging[x,y]-1}{\max(kriging-1)} \tag{10}$$

The quality of the interpolation depends on the number of measurement points and even for kriging the number of measurement points as derived from the human-subject experiment is low. The effect of that is that the fitted variogram model is less stable. To increase the number of measurement points and thus the quality of the resulting field, additional measurement points were created based on the properties that the analysis described in section 3.2 revealed. For the cardinal directions where direction plays a the primary role the measurement locations were mirrored across the cardinal direction's primary axis[17] (fig. 24). With "near" the analysis showed that angle played no significant role, it was thus possible to mirror the measurement locations across both axis, effectively quadrupling the number of measurement locations (fig. 25) and making the resulting field more stable.

[17] For north and south this is the vertical axis, for east and west the horizontal axis

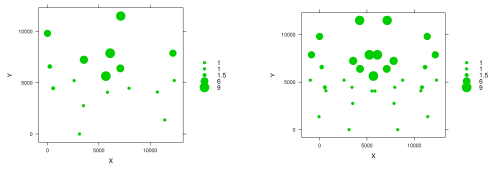

Fig. 24 The original measurement points for "north" on the left and the mirrored, duplicated measurement points on the right.

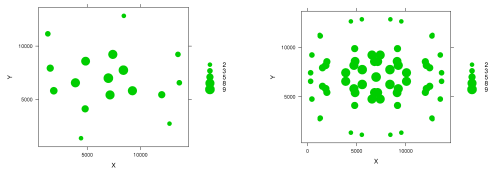

Fig. 25 The original measurement points for "north" on the left and the quadrupled set of measurement points that is used in the final implementation on the right

8.3 Accessing the field values

The `read` operation provides access to the vague-field's applicability values. It translates between the external, unbounded, continuous representation and the internal, discrete, bounded field-value matrix. The translation from the external to the internal representation is performed using the internal and external anchor locations. The offsets of the external x and y coordinates relative to the external anchor location are calculated and then using the field's scale value transformed into internal offset coordinates. These internal offset coordinates are then added to the internal anchor's coordinates to determine the internal coordinates. The internal coordinates are then used to read a value from the field matrix, which is returned.

8.4 Combining fields

The field-combination calculation (eq. 1) is performed every time the combined field is accessed. While this is a computationally expensive approach, it has the advantage that fields of any scale can be combined without having to align their internal matrix representations, as the fields are accessed through the `read` function and can thus be treated as continuous and scale-free.

The normalisation factor nf is defined as the maximum combined field value and is calculated by placing a virtual grid over all fields, calculating the value at each grid point and taking the maximum of these values. One problem with this approach is that if the source fields' cells overlap as shown in figure 26 then none of the maximum measure-

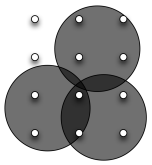

Fig. 26 Three fields that overlap at their boundaries and where none of the measurement points used to calculate the field maximum (small white circles) measure the actual maximum where the three fields overlap.

Fig. 27 The three vectors that define the direction the control point will move. The dashed line represents the field's vector, the dotted line is the contraction field vector and the small final arrow is the internal energy vector. In the left-hand case the control point will be moved one grid-cell in the direction indicated, while in the right-hand case the control point will not move, as a local minimum has been found for that control point.

ment points actually measure the combined maximum. This means that if the combined field is read at a location that would produce the actual maximum then the calculated value would be larger than the normalisation factor and the resulting value would be larger than 1, which is not allowed. To avoid this if the combined value is larger than the normalisation factor, then a value of 1 is returned, regardless of what the actual measurement value is. While this may seem to skew the data, if all the source fields are continuous, as is usually the case, then the difference between the calculated and the actual maximum is very small and can be disregarded.

8.5 Crisping the vague-field

The `crisp` operation is used to transform the continuous vague-field into a crisp polygon for integration with existing GI systems and algorithms and is based on active contours.

Active contours The concept of active contours was introduced by Kass et al (1988) as a method of finding boundaries in image data, but have also been used in GIS for various purposes (Burghardt (2005)), Steiniger and Meier (2004)), and Horvath et al (2009)). They are defined as controlled continuity splines (Terzopoulos (1986)) upon which image and external forces act to move them into the desired shape. In the original method the energies acting upon the active contour are defined as in equation 11, consisting of an internal energy, the image energy and external constraint energy, which the active contour then tries to minimise.

$$\begin{aligned} E^*_{snake} &= \int_0^1 E_{snake}(v(s))ds \\ &= \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))ds \end{aligned} \quad (11)$$

The internal energy acts to maintain the active contour's shape, image energy can be defined via the image intensity (Kass et al (1988)), image gradient (Lam and Yan (1994)), or via more complex methods (Xie and Mirmehdi (2006)), and the external energy defines constraints that the active contour needs to observe that are not directly defined by the active contour itself or the image data. An iterative method on a grid is used to move the active contour's control points to their final solution. For each control point the minimum energy neighbour is calculated and the control point moved there immediately. The energy calculation for the next control point will thus take into account the updated position of the previous control point. This is repeated until the active-contour's final shape is found and means that the active contour will achieve a locally minimal solution, but not necessarily a globally minimal solution. Due to this iterative way of moving, active contours are often also referred to as snakes, as they seem to slither across their processing space (fig. 30).

Crisping fields with active contours To enable the use of active contours in creating a crisp representation of the vague-field, a slightly modified energy function is used (eq. 12). The first two energies, internal and field, are similar to the internal and image energies as defined earlier. The contract energy is an external energy that pulls the active contour towards the centre of the field.

$$E_{snake} = \alpha \cdot E_{int} + \beta \cdot E_{field} + \gamma \cdot E_{contract} \quad (12)$$

Each energy is defined as a vector field, with the direction of each cell's vector defining the direction in which an active contour control point at that location would be pushed (fig. 27). The length of the cell's vector defines how strongly the control point is being pushed in the specified direction, thus the energy function (eq. 12) can be implemented as a simple vector addition, with the final vector defining the direction the control point will move.

In this framework the internal energy (eq. 13) is defined as the vector from the control point (p_i) to the centre-point between the preceeding (p_{i-1}) and following (p_{i+1}) control points (fig. 28). This definition ensures that the control points always move so as to create a snake where the control points are evenly spread, since the further a control point moves towards its predecessor and further away from is successor, the stronger it will be pulled towards the successor.

$$E_{int} = \left(v_{prev} + \frac{v_{next} - v_{prev}}{2} \right) - v_{current} \quad (13)$$

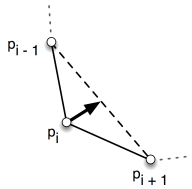

Fig. 28 The internal energy is defined as the vector from the current point (p_i) to the half-way point between the previous (p_{i-1}) and the next control point (p_{i+1}).

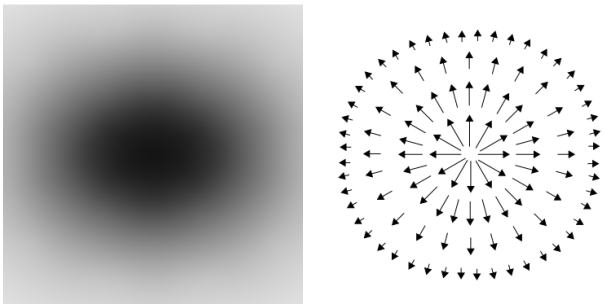

Fig. 29 The vague field for "near" and a simplified representation of its gradient field.

The scalar vague-field is transformed into the required vector representation by applying the gradient operator. The gradient operator defines each cell's vector so that it points in the direction of the neighbouring cell with the lowest scalar value. The length of the vector is determined by the value of the current cell, unless the minimum is equal to the cell's value in which case the vector's length is set to 0 as the cell is a local minimum (fig. 29).

The contraction energy is used to define how far the active contour will contract. It is defined as a constant vector field of the same extent as the vague-field that is being crisped, with each cell's vector pointing towards the centre of the vague-field and of length 1. The centre of the vague-field is defined as the centroid of all cells with the maximum value. This guarantees that the active contour will contract towards the strongest part of the field.

In the active contour energy function (eq. 12) each component energy has a weight attached to it, to define their relative influences on the total energy. The weights have been tuned experimentally and in the results shown in this section are $\alpha = 0.2745$, $\beta = 1$, $\gamma = 0.4314$. The weights were chosen so as to create crisp polygons that had extents that roughly matched the angles and distances observed in the initial Geograph experiment (Hall and Jones (2008)). The snake is initialised and terminated so as to minimise the number of iterations it has to run through, while guaranteeing a valid result. Details of the methods used to enable this can be found in Hall and Jones (2009).

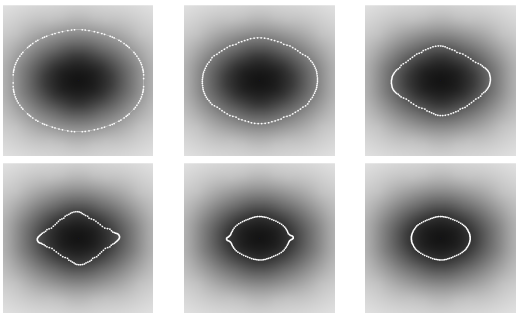

Fig. 30 The active contour moving from the initial location to its final solution on a field for "near". To illustrate the principle, the snake was initialised at $\alpha = 0.4$.