"Where the city sits?"
Revealing Geospatial Semantics in Text Descriptions

Eleni Tomai, Marinos Kavouras
Cartography Laboratory, NTUA
Athens, GREECE
etomai@central.ntua.gr, mkav@survey.ntua.gr
http://ontogeo.ntua.gr

SUMMARY

This paper seeks to reveal the semantics of geospatial concepts as it can be elicited from text descriptions. So far, such an approach has dealt with the analysis of existent repositories of geographic information. Taking up with the study of the semantics of geospatial concepts, we turn to the analysis of the linguistic expressions of space and consequently of the "things" it includes. More specifically, we identify linguistic terms that people use when they locate themselves in space or when they move around it. In addition, we examine how people tend to portray objects positioned in space and which descriptions they use to communicate with each other about these objects. From this information, inferences can be drawn on what constitutes the semantics of spatial objects and how this is connected to natural language. What is more, the generation of mental maps has employed spatial descriptions due to their expressive power and use of natural language. Consequently, modern cartography has a lot to gain from the semantic analysis of text descriptions of space.

KEYWORDS: *Linguistic expressions, descriptions, geospatial objects, semantics*

INTRODUCTION

Expressions of spatial reference and geographic concepts are common to all natural languages. The reason for that is that humans are spatially referenced "objects", that is, they are forced to interact with space in every aspect of their existence. Subsequently, the wealth of their natural languages in linguistic expressions about space is fully justified. What is more, natural languages have deployed figurative expressions when describing spatial objects, such as metaphors with great expressive power. The ground for this approach is that the analysis of spatial linguistic expressions in everyday language can reveal the semantics of geospatial objects and therefore account for their particular characteristics.

Prior research on the subject has focussed on the study of definitions of geographic concepts met in geographic data repositories and terminological databases with the purpose to extract information as the basis of integrating or generating ontologies. Herein, we go a step further in this endeavour by analysing free text, such as a small but indicative piece of travel guide texts, to identify all those linguistic expressions used either to describe geospatial objects or to give instructions on how to reach them. The rationale behind this undertaking is to draw inferences on what constitutes the semantics of space and the "things" it includes and how this is associated with the use of language. Also, because the selected corpus is constituted by texts found on the Web, the application of such an analysis will facilitate the understanding of the semantics and sharing of the information represented by the data.

FORMER RESEARCH

This section gives an account of what is been achieved so far. Prior research and its results have been presented comprehensively elsewhere, for that reason; herein we give only a brief account of the approach and its outcomes.

The Approach and Its Results

As presented elsewhere (Kokla & Kavouras, 2002, Kavouras et al., 2003), our approach consisted in analysing the definitions of geographic concepts in terms of a set of semantic relations and semantic properties. The analysis takes place by the identification of lexical and syntactic patterns found in definitions, which constitute a well-structured text. The use and outcomes of the approach have also been demonstrated elsewhere in detail, so here a brief description will be given. The analysis has shown the following:

- Several semantic properties and semantic relations that are present in the definitions of geographic concepts (Kokla & Kavouras 2002) have been pinpointed using Natural Language Understanding.
- Subsequently, by comparing existent geographic information repositories, numerous inconsistencies have been revealed (Kavouras et al., 2003)
- The technique was further applied for integrating these distinct repositories (Kokla & Kavouras, 2002), and pursued in providing a framework for semantic integration (Kavouras, 2003).

Other attempts to understand the semantics of definitions of spatial concepts and use this in the integration of ontologies can be found in Hakimpour and Timpf (2002) and Kuhn (2002). However, an existing plethora of texts describing objects, having spatial reference, challenged us to pursue this kind of research in revealing the semantics of spatial language and therefore, of spatial concepts, in analysing this kind of corpora.

PROPOSED APPROACH – FROM DEFINITIONS TO LINGUISTIC EXPRESSIONS

Hitherto, we have focussed our research of Natural Language Understanding to the analysis of the geographic concepts definitions. Currently, we are proceeding into revealing the semantics of descriptions of geospatial objects. The case study that will be presented here includes the analysis of texts, which can be found in travel guides. Their selection was made because they present an ample repository of spatially referenced linguistic information, they are widely used by our community, and they include descriptions of space using figurative language apart for literal accounts of spatial expressions.

Case Study Texts

The texts to be analysed here are excerpts of tourist guides found on the Web. The selected passages contain descriptions of European regions and one city. All four analysed texts were taken from http://www.fodors.com/miniguides/ and the selected "destinations" are: Brittany, Andalusia, Peloponnese, and Edinburgh.

ANALYSIS

The scrutiny of the linguistic expressions used in the selected corpora, which describe geospatial objects and phenomena revealed the following about the semantics of *spatial language* as termed hereinafter. The major axis of the study covers the notions of location and motion. Another important constituent of the texts is the extended usage of descriptive and figurative language. The following paragraphs report the findings.

Description of Location

Locating spatial concepts involves the use of special linguistic terms such as the category of prepositions, which accompany the locative verb. Due to the variety and many times the complexity of topological

relations in the geographic space, language turns to more complicated lexical items than prepositions, which make use of reference specifications. The above notions are portrayed in the following examples as found in our case study corpora.

1. And scattered *over* the mossy hillsides *stand* Stonehenge-like dolmen and menhirs …
2. … this southernmost part of Greece, …, *situated in* equally varied and beautiful scenery …
3. And, *at* the very tip of continental Europe *dangles* the Mani peninsula.
4. … a mountain of bright green … *rearing up behind* the spires of the Old Town.
5. And high on the *southern* slopes of the Sierra Nevada, (*are*) the villages of the Alpujarra

Examples 1 and 2 reflect the topological relation between geospatial objects at its simplest form, in which, following Talmy's distinction between the figure and the ground (Talmy, 2000), we can identify which geospatial object is located in reference to the other. Examples 3 to 5 on the other hand, exemplify differently the topological relation, using "coordinate systems", because they try to inter-locate disjoint geospatial objects. In literature, these systems are called *frames of reference* and are distinguished in the intrinsic, relative and absolute frame of reference (Levinson, 2003). More specifically, example 3 uses the intrinsic frame of reference partitioning the ground object into parts to locate the figure object in reference to these parts. Example 4 applies the relative frame of reference to locate a geographic feature in reference to the communicant's (speaker, viewer, writer) view of the ground object. Example 5 makes use of the absolute or geocentric frame of reference.

Description of Motion

In literature, the distinction between path-manner languages, first discussed by Talmy in 1985 (Talmy, 2000), is made to distinguish over different grammatical and syntactic patterns that express motion. Because our test-bed is English, which is a manner language, further analysis of the distinction is not undertaken herein. Motion, in the case of manner languages, is coded by the motion verb and the prepositional phrase, which follows it, gives the path information (Papafragou et al, 2002). This is not to be taken to mean that path verbs do not exist in manner languages as well, rather that manner languages favour the use of manner grammatical and syntactic patterns contrary to path ones. The examples in the following paragraphs make the path-manner distinction obvious.

"Motion" of geographic objects

Motion of geospatial objects - according to Talmy's (2000) terminology: fictive motion - can refer to as a location in space but using a "vector" which is exemplified by the *source* and *goal* of the motion. The source and goal constitute the ground objects. Examples are given below.

6. The Northern Peloponnese consists of the Argive peninsula, which juts into the Aegean… and *continues westward* past the isthmus and *along* the Gulf of Corinth *to* Patras and the Adriatic coast.
7. The Guadalquivir, …, *runs through* the entire region.

Example 6 traces the moving figure using verbs of motion such as *continue*, with multiple ground objects instead of one, in order to set the beginning and ending of "motion". While in example 7, "motion" description of the figure makes use only of one ground object as landmark without coding the origin and termination of the motion (i.e., watercourse), which seems to be irrelevant in the particular context.

Expressing motion events in space

The dynamic displacement (or its possibility) of other objects or agents in space is coded in the given texts as well. Take a look at the following examples.

8. Wherever you *wander in* Brittany -- *along* jagged coastal cliffs, *through* cobbled seaport streets, *into* burnished oak cider pubs …
9. In the Southern Peloponnese, those who *penetrate* the forbidding mountains of the Taygettus range …

Example 8 reflects the manner dimension of motion, with the use of verb *wander* to mean *to go at no set pace*, and the path is encoded by the prepositional phrases that follow. On the other hand, the next

example reflects the path dimension of motion as encoded in language through the verb alone, expressed by the verb *penetrate: to pass in or through* (both definitions taken from WordNet).

The Use of Descriptive Language

Besides the two depicted notions of location and motion, the analysis has also identified the existence of descriptive language in the case study texts. This kind of language plays an attributive role. The use of adjectives and adverbs denote the most "salient" features, in the context and for the purpose of the text itself, of the described landscape. The information, given this way, ranges from *size* to *shape* and from *use* to *state* (refer to semantic properties as in Tomai & Kavouras, 2003) aiming to provide a - concrete yet made somewhat more appealing, but nevertheless true - mental image of the landscape. Apart from the mentioned linguistic categories, descriptive language of the given texts includes metaphors, similes, and personifications: elements that belong to figurative language.

10. Wherever you wander in Brittany -- along *jagged coastal* cliffs, through *cobbled seaport* streets, into *burnished oak cider* pubs …
11. Typical Andalusian architecture - *brilliant-white* villages with *narrow*, *shady* streets; *thick-walled* houses clustered around cool, private patios; *whitewashed* facades with modest, grilled windows - comes from …
12. … an eastward-grinding glacier encountered the tough basalt core of the volcano and swept around it, scouring steep cliffs and leaving a trail of matter, *like* the tail of a comet.
13. Its (Edinburgh) *shape* was becoming clearer: *like a fish with its head at the castle*, *its backbone running down the ridge, and its ribs leading briefly off on either side*. *The backbone gradually became the continuous thoroughfare*, … and *the ribs became the closes* (alleyways)...
14. Edinburgh Castle *watches* over Scotland's capital city, *frowning down* on Princes Street.

Examples 10 and 11 are representative of expressing the attributive role of descriptive language, which makes the representation of the environment possible through mental images. Example 12 uses both metaphor and simile to generate an image. In accord, example 13 uses a simile to express a property such as shape and recognizes where and how parts of the geographic object (in this case the city of Edinburgh) are located. Then metonymy is applied to refer to the current state of part of the city. Finally, example 14 uses personification to express location, and orientation that is, spatially referenced attributes of the geographic objects.

ACCOUNT OF THE FINDINGS

The previous examples have exposed ways of describing space and the objects that constitute it. At this point, it is necessary to illustrate the worthiness of such an analysis in the study of space itself. Such an account of the findings will range along the following dimensions: the semantics of location and motion, the cognitive aspect of figurative and descriptive language, the value of spatial descriptions in knowledge acquisition.

Semantics of Location and Motion

As made clear from the analysis, location, and motion (either "motion" of geographic objects or motion of agents or other objects in geographic space) are central in the description of space and spatial objects at least in the case examined. Even though these two notions are frequent in language, natural languages exhibit variation in expressing them, as seen clearly by the distinction of path-manner languages. However, location and motion are semantically prominent in human reasoning about space across cultures and languages and as proved by cross-linguistic studies, word differences do not fully reflect differences in conceptual representations (Gennari et al., 2002, Papafragou et al., 2002). Put it bluntly, location and motion semantics are expressed differently cross-linguistically, nevertheless grasped the same way.

Cognition and Descriptive/Figurative Language

Descriptive language aids in the direction of "sketching" images in the mind that is why it is widely used in the examined herein context. Figurative language on the other hand, which has been long thought as a *linguistic embellishment*, has recently been given a cognitive value (Lakoff, 1987). As seen from the relevant examples, these "figures of speech" are used both to reveal and illustrate characteristics of spatial objects. The pragmatics dimension (Papafragou, 1996) of such elements of speech provides them with their cognitive aspect, so that by using them representations of space can be generated.

Spatial Descriptions and Knowledge Acquisition

Describing space in one thing, to convey a representation of it through text descriptions is another. Knowledge acquisition of spatial concepts is a complex task, which involves more than descriptions (Raubal & Winter, 2002). Mental representations of spatial language have proved to be a very powerful tool for *computer-assisted problem-solving performance abilities* (Hubona et al., 1998).

Nevertheless, the analysis of corpora that use a wide variety of linguistic expressions, and syntactic structure may provide a device for revealing "prominent" geographic objects, their "salient" properties (either spatially referenced such as location, or not, such as shape), and for establishing the semantic typology for the language elements capable of performing this task.

MERIT OF THE APPROACH

This paper sought to reveal how text descriptions encode geospatial semantics. In the case of the herein selected corpus, location and motion proved to be central notions for informing and orientating the readers, when describing "destinations". In addition, descriptive language is present throughout these texts as a mean to generate mental images of the described places.

The analysis can be used in providing a typology of the "spatial language" that humans use to communicate about space and its objects. Utterances and written descriptions of locating things in space or locating and moving oneself around it, deploys different natural language resources such as prepositional phrases, and motion verbs. On the other hand, in our attempt to convey an image of a (geographic) site we turn to use "sophisticated" language such as similes, hyperboles, and metaphors to refer to features of geospatial objects, such as size, shape or topology. These linguistic elements tend to differ widely over natural language cultures and their exhaustive analysis will give a clear account of how humans reason about space and express it through language.

What is more, due to the fact that the specific corpus is found on the Web, the included "spatial language" is worth identifying and analysing, for many attempts of formalizing data, currently involve data found on the Web. Consequently, establishing a semantic typology of the language used in such an environment is consistent with major trends of the research for formalizing the content of the Web. In the field of geographic information science, this kind of research is crucial in order to grasp aspects of geospatial concepts when people pen their descriptions or even indite them.

FURTHER WORK

The paper examined which linguistic terms articulate geospatial objects in terms of their attributes. Moreover, it attempted to analyse which syntactic patterns describe location and motion in space and how these contribute to the semantics of these notions. The analysis presented here is far from an inclusive account of all these elements that constitute the so-called spatial language, however we managed to isolate its major dimensions.

Further work is needed in terms of how natural language aids in the direction of cognition and knowledge acquisition of space. Cross-linguistic studies should be conducted to give a complete account and understanding of how natural languages represent spatial semantics. In this respect, our motivation is to

carry out similar research for Modern Greek as well as English with the purpose of revealing how different linguistic cultures locate themselves or move around space from text descriptions.

BIBLIOGRAPHY

Fodor's Travel Publications, http://www.fodors.com/miniguides/

Gennari S. P., Sloman S. A., Malt B. C., & Fitch W. T., Motion events in language and cognition. In Cognition, 83: 49-79, 2002.

Hakimpour F., Timpf S., A Step towards Geodata Integration using Formal Ontologies. In M. Ruiz, M. Gould, J. Ramon (eds.). Proceedings of the 5th AGILE Conference on Geographic Information Science, Universitat de les Illes Balears: 55-60, 2002.

Hubona G. S., Everett S., Marsh E. & Wauchope K., Mental representations of spatial language in International Journal of Human–Computer Studies, 48: 705–728, 1998.

Kavouras M., A Unified Ontological Framework for Semantic Integration, International Workshop on Next Generation Geospatial Information, Cambridge (Boston), Massachusetts, USA, 2003, http://www.ontogeo.ntua.gr/publications/english_pub.htm.

Kavouras M., Kokla M. & Tomai E., Determination, Visualization and Interpretation of Semantic Similarity among Different Geographic Ontologies. In M. Gould, R. Laurini, S. Coulondre, (eds.). Proceedings of the 6th AGILE Conference on Geographic Information Science. Presses Polytechniques et Universitaires Romandes: 51-56, 2003.

Kokla M. & Kavouras M., Extracting Latent Semantic Relations from Definitions to Disambiguate Geographic Ontologies. In G. Zavala (ed.). GIScience 2002 Abstracts. University of California Regents: 87-90, 2002.

Kuhn W., Modelling the Semantics of Geographic Categories through Conceptual Integration". In M. Egenhofer D.M. Mark (eds.), Geographic Information Science. Lecture Notes in Computer Science. Springer Berlin, Vol. 2478: 108-118, 2002.

Lakoff G., 1990 Women, Fire, and Dangerous Things: What Categories Reveal About the Mind University of Chicago Press (Trd); Reprint edition. ISBN: 0226468046, pp. 614.

Levinson S., Spatial language. In L. Nadel (ed.). Encyclopaedia of Cognitive Science, London: Nature Publishing Group, 4: 131-137, 2003.

Papafragou A., Massey C., & Gleitman L, Shake, Rattle, 'n' Roll: The Representation of Motion in Language and Cognition. In Cognition, 84: 189-219, 2002.

Papafragou A., Figurative Language and the Semantics-pragmatics Distinction. In Language and Literature, 5: 179-93, 1996.

Raubal M, & Winter S., Enriching Wayfinding Instructions with Local Landmarks In M. J. Egenhofer, D. M. Mark, (eds.), Geographic Information Science. Lecture Notes in Computer Science. Springer Berlin, Vol. 2478: 243-259, 2002.

Talmy L., 2000 Toward a Cognitive Semantics Volume 1, Concept Structuring Systems. Cambridge, MA: MIT Press. ISBN 0262201208, pp. 480.

Tomai E. & Kavouras M., Pivotal Issues in Designing Geographic Ontologies, COSIT-03 - Workshop on Fundamental Issues in Spatial and Geographic Ontologies, Ittingen, Switzerland, 2003, http://ontogeo.ntua.gr/publications/tomai&kavouras_workshopCOSIT03.pdf.

WordNet 1.7.1 - a Lexical Database for the English Language, Cognitive Science Laboratory, Princeton University, http://www.cogsci.princeton.edu/~wn/