

Methodologies and Resources for Processing Spatial Language

31 May 2008

Graham Katz (Georgetown)

Inderjeet Mani (MITRE)

Thora Tenbrink (Bremen)

ABSTRACTS

WORKSHOP PROGRAMME

PROGRAM

9.00 Welcome & Introduction

9.10 - 9.50 Invited Talk:

The Long Road from Spatial Language to Geospatial Information, and the Even Longer Road Back: The Role of Ontological Heterogeneity

John Bateman (University of Bremen)

9.50 - 10.30 Placename Ambiguity Resolution

Geoffrey Andogah, Gosse Bouma, John Nerbonne, Elwin Koster (University of Groningen)

BREAK

11.00 - 11:40 Annotating Natural Language Geographic References

Inderjeet Mani, Janet Hitzeman and Cheryl Clark (MITRE)

11:40 - 12:20 Spatial Entities are Temporal Entities Too: The Case of Motion Verbs

Nicholas Asher, Philippe Muller, and Mauro Gaio (IRIT/CNRS)

12:20 - 12.40 Reusable Grammatical Resources for Spatial Language Processing (with **demo**)

Robert J. Ross (University of Bremen)

12.40 - 13.00 Realtime Path Descriptions Grounded with GPS Tracks: A Preliminary Report (with **demo**)

Nate Blaylock and James Allen (IHMC/University of Florida)

LUNCH

2:30 - 3:10 Building a Parallel Spatio-Temporal Data-Text Corpus for Summary Generation

Ross Turner, Somayajulu Sripada, Ehud Reiter and Ian P Davy (University of Aberdeen and AMI)

3:10 - 3:50 A Field Based Representation for Vague Areas Defined by Spatial Prepositions

Mark M. Hall and Christopher B. Jones (Cardiff University)

3:50 - 4:00 Concluding Remarks/Discussion

W-12 Methodologies and Resources for Processing Spatial Language

Saturday, May 31, 9:00

The Long Road from Spatial Language to Geospatial Information, and the Even Longer Road Back: the Role of Ontological Heterogeneity

John Bateman; University of Bremen, Germany

In this talk I present from an ontological perspective some of our ongoing work on the situated interpretation of natural language concerning space. Our long term goal is to construct systems which are capable of dealing with the full range of flexibility observed in natural language concerning space rather than adopting artificial restrictions for the purposes of particular applications or tasks. Our treatments have been particularly concerned with capturing spatial relations: that is, the semantics of utterances where distinct entities are brought into a spatial configuration where more or less specifically specified relations hold between them. A major challenge in dealing with natural spatial language usage is the observed flexibility of the language used: there is no simple strategy for assigning spatial terms to, for example, geometric configurations, field descriptions and so on. The conclusion we draw from extensive analyses in the area is that it is necessary to cleanly separate the linguistic semantics of space from the non-linguistic, situation-specific interpretation of space. In particular, we consider language itself to contribute an ontology-like organisation of the spatial world and so pursue this as an application of ontological engineering.

Placename Ambiguity Resolution

Geoffrey Andogah, Gosse Bouma, John Nerbonne, Elwin Koster; University of Groningen, Netherlands

It is common for placenames to reference other named entities (e.g., names of people, names of organizations, etc.) and to be used as vocabulary words (e.g., city of Split). Apart from reference ambiguity, placenames are faced with the problem of referent ambiguity (i.e., a placename referring to multiple places). Many places are also referred to by multiple names (e.g., Netherlands vs. Holland). In this paper we describe an approach to place ambiguity resolution in text, i.e., place reference resolution, resolution of a document's geographical scope and placename referent resolution. The approach is composed of three components: (1) geographical tagger, (2) geographical scope resolver and (3) placename referent resolver.

Annotating Natural Language Geographic References

Inderjeet Mani, Janet Hitzeman and Cheryl Clark; MITRE, USA

SpatialML is an annotation scheme for marking up references to places in natural language. It covers both named and nominal references to places, grounding them where possible with geo-coordinates, including both relative and absolute locations, and characterizes relationships among places in terms of a region calculus. A freely available annotation editor has been developed for SpatialML, along with three annotated corpora, including a corpus of annotated documents released by the Linguistic Data Consortium. Inter-annotator agreement on SpatialML extents is 77.0 F-measure on that corpus, but 92.3 F-measure on another (ProMED) corpus. The paper discusses a number of issues affecting inter-annotator agreement.

Spatial Entities are Temporal Entities Too: The Case of Motion Verbs

Nicholas Asher, Philippe Muller, and Mauro Gaio; IRIT/CNRS, France

We argue that spatio-temporal primitives are crucial in giving a full view of the spatial and temporal structure of texts. We think that temporal and spatial structure are projections of a more complex and more complete spatio-temporal structure. We will make our case based on the analysis of movement verbs, showing how they contribute in an important way to both temporal and spatial structure within discourse. Our analysis of movement verbs is based on a detailed lexical semantic analysis of a wide class of verbs in French. We give some ideas for how this lexical semantics when coupled with an analysis of how clauses involving these expressions are related to each other within a discourse using rhetorical relations can aid in determining the spatio-temporal structure of the text. We apply our approach to descriptions of climbing cliffs as well as descriptions of walking tours in the Pyrenees and descriptions of itineraries in Toulouse. We think that this provides sufficient justification for including movement verbs and spatio-temporal information in general within the specification of a SpatialML or rather its fusion with TimeML.

Reusable Grammatical Resources for Spatial Language Processing

Robert J. Ross; University of Bremen, Germany

While modularity and resource reuse are key to rapid development in the software engineering community, the emphasis on re-use has not been equalled in components which provide linguistic interfaces to spatial applications. As a step towards reducing this problem, we demonstrate open-source grammatical resources which we are developing for applications which require a clear and controllable spatial language interface. The grammars, based on Steedman's Combinatorial Categorical Grammar (CCG), can be used for both language analysis and production in the German and English languages, and have been fused with a rich spatial semantics based on a well-defined

linguistic ontology. We demonstrate the use of the grammars themselves through the publicly available OpenCCG tool, and also illustrate their use in an implemented dialogue based spatial language processing application for navigating agents.

Realtime Path Descriptions Grounded with GPS Tracks: A Preliminary Report

Nate Blaylock and James Allen; IHMC/University of Florida, USA

In this paper, we present a novel method for data collection which produces aligned real-time speech recordings of path descriptions and the corresponding GPS track of that path in the real world. We give a preliminary report on the pilot corpus we have gathered using this method and our initial annotation plans for references to location, orientation and movement within the speech. Using the GPS track of the real path and a GIS database, we plan to annotate spatial references in the corpus with the "ground truth" of objects in the GIS database, or lat/lon coordinates corresponding to referent object or location. This annotated data will provide a set of natural language descriptions of paths, locations, orientation and movement which can be used for training/testing algorithms for understanding spatial language situated in the real world and aided by a GIS database. We also describe an initial annotation tool we are building that uses Google Earth for visualizing and annotating the corpus.

Building a Parallel Spatio-Temporal Data-Text Corpus for Summary Generation

Ross Turner, Somayajulu Sripada, Ehud Reiter and Ian P Davy; University of Aberdeen and AMI, UK

We describe a corpus of naturally occurring road ice weather forecasts and the associated weather prediction data they are based upon. We also show how observations from an analysis of this corpus have been applied to build a prototype Natural Language Generation (NLG) system for producing road ice forecasts. While this corpus occurs in a narrow domain, it has much wider applicability due to the nature of its spatial descriptions, whose primary communicative goal is to describe the interaction between meteorological parameters and geographic features.

A Field Based Representation for Vague Areas Defined by Spatial Prepositions

Mark M. Hall, Christopher B. Jones; Cardiff University, UK

Natural language is one of the primary means of communicating spatial information, but existing geographic information retrieval (GIR) facilities are weak in this respect. One of the major challenges in automated interpretation of spatial natural language is how to model the regions implied by spatial expressions. This paper presents a field-based model for representing the vague regions defined by spatial language, including a method for defining field values from existing spatial language data sources. To interface this new vague field model with existing GIR systems and methods, an algorithm for extracting crisp boundaries from the field representation is also presented.