Semantics of Landscape: Providing Information About the Landscape via The Semantic Web

David M. Mark
NCGIA-Buffalo
Semantics

- Semantics is all about meaning
- Semantic interoperability is about the preservation of meaning as information is exchanged among software, databases, and people
Words

• Words in natural language play a critical role in the transmission of meaning from one person to another...

• and in the sharing of concepts among individuals within speech communities
Ontologies

• Recent developments in ontologies for knowledge representation also have words and their meanings playing a central role

• Such ontologies mainly formalize the definitions of and relations among words (terms) for a specific domain
Conceptualizations

• If an ontology is an explicit specification of a conceptualization, where do the conceptualizations come from??
The Key Role of Words

- The key role of words and language in semantics is recognized in the ubiquitous semiotic triangle, which relates:
  - signs (symbols, including words)
  - concepts (mental ideas), and
  - referents (in the real world)
"The Meaning Triangle"

Semantics of what?

• In contrast to the Münster group, I am trying to capture the meaning of natural language expressions!

• Formalizations of the common-sense meanings of words are needed for building ontologies.
Semantics of Landscape

• The natural landscape is a very important domain for human activities, and is particularly important in geospatial intelligence

• Representation and exchange of information about the landscape in word-centered ontologies such as those alluded to above, for delivery through the Semantic Web, presents a number of significant research challenges that are rare or absent for other domains
Some Major Issues

• Semantics usually deals with objects but geospatial information is often fields or images
• Classifications of landscape elements vary across cultures and languages
• Delimitation of landscape objects may be dependent on categorizations
• The above issues interact!
Semantics usually deals with objects but geospatial information is often fields or images

- While the semantic web is word-centered, the great majority of existing geospatial information about natural landscapes is in the form of fields (digital elevation models) or images (mainly from remote sensing).
- Serving of field-based or image-based geospatial information via the semantic web will require robust computational methods for identifying and delimiting landscape features present implicitly in those fields and images.
Classifications of landscape elements vary across cultures and languages

• For inorganic natural domains such as landforms and water bodies, natural variation usually forms a continuum, and categories for landscape elements thus may have a somewhat arbitrary component, contingent on landscape variation, lifestyle of the people, and linguistic and cultural history
Landscape Objects are Dependent on Classifications

• For many domains, the entities dealt with are bona fide objects that exist as objects independently of their definitions
• In contrast, many landscape elements are not bona fide objects, but are parts of a continuous land surface
• Thus different definitions of landscape element types may lead to a different subdivision of the land into objects
The above issues interact!

- The same landscape, represented by the same field-like data, may produce quite different delimited objects and object classes according to the definitions of words in different languages.
- Higher-level feature-extraction methods probably will need to be language-specific.
- A word-based ontology of natural landscapes developed on the basis of one natural language (e.g. English) probably will be inadequate for other languages.
Prognosis

• The extent of these issues is not yet known, but several well-documented examples suggest that they will be a substantial impediments to multilingual semantic interoperability for landscape data using the word-centered approaches now common for knowledge-representation ontologies and for the Semantic Web.
Conceptualizations

• Research is needed to document the terms and associated concepts that everyday people use to structure and talk about the common-sense world.

• Ethnographic methods are needed to compile these terms and their definitions, which only then can be formalized à la Gruber!
Conceptualizations of Landscape

• We do not yet have a complete basis for doing this even for English, although standards such as SDTS and DIGEST, and knowledge bases such as WordNet and CYC, are good starts
Linguistic Relativism

• Different languages *lexicalize* different concepts
• Therefore, an ontology based on the meanings of words in one language *generally will not be adequate for representing concepts for other languages*
Landscape ‘Concepts’ That Have No Single Word in English

• “A spot of level ground in the mountains, surrounded by ridges”
• “A type of hollow in a sandhill, used as camping place, especially in cold weather”
• “A canyon wall receiving sunlight”
• “An ‘island’ of land completely surrounded by one or more younger lava flows”
• “An island of grassland left unburnt after a surrounding wildfire”
Example 1: **Categories for Water Bodies**
How Many Kinds of Water Bodies?
Conceptual Model for Water Bodies

• Kinds of bodies of water may be distinguished along several dimensions:
  – Size
  – Flowing or still
  – Salt or fresh water
  – Other aspects of water quality
  – Origin
  – Seasonality of water
  – Seasonality of flow
  – ...

• Different languages or cultures may give different weights to these factors
French and English Categorize Standing Water Bodies Differently
Example 1, Summary

- All *ponds* are *étangs*, but not all *étangs* are *ponds*
- The main defining character of an *étang* is that it has no surface outlet; English does not use that characteristic in category definitions
- If we were trying to translate an instance of the word *étang* from French to English, we would need to know what kind of water body we were referring to before we could choose the correct English word
- Word for word translation is not possible without added information because the *concepts* do not line up!
Example 2: “Hills” in Yindjibarndi
marnda
marnda

bargu

burbaa
mountain

marnda

bargu

burbaa

hill
Scale difference in the terminological break
Scale difference in the terminological break
A Many-to-Many relation

Scale difference in the terminological break
Summary: Marnda, Hill, etc.

• There is a many-to-many relation between {mountain, hill, ridge, range} and {marnda, bargu, burbaa}

• It would not be possible to correctly translate “marnda” into English or “hill” into Yindjibarndi without knowing the size of the landform being referred to

• This appears to be a different way to subdivide instances into categories, but based on similar principles
Example3: Watercourses in Yindjibarndi

- There are no permanent or even seasonal rivers or creeks in Yindjibarndi country. Larger watercourses have running water in them only after cyclones.
- Permanent sources of water include permanent pools along the channels of the Fortescue River, as well as some permanent small springs, and soaks, where water can be obtained by digging.
Watercourses and their beds in Yindjibarndi and English

• According to the dictionaries, Yindjibarndi has the following words:
  • Wundu: “river(bed), gorge” (Anderson)
  • Yijirdi: “small stream, rivulet” (Anderson)
  • Manggurdu: “Fortescue River” (Anderson)
  • Mangkurdu: “flood” (Von Brandenstein)
Creeks and Creek-beds in Yindjibarndi and English

• We currently believe that the linguists got confused about what was being named
• There are no permanently-flowing rivers or creeks in Yindjibarndi country
• Meanings of Yindjibarndi terms were (apparently) shoe-horned into an English-language, permanent stream ontology
English ‘ontology’ and terminology

“creek”
or
“creek bed”

“creek”
“creek bed”
Yindjibarndi ‘ontology’ and terminology

“wundu”

“yijirdi[?]”

“wundu?”
Which is the feature, the water, or the creek bed?
# Rivers and Riverbeds

<table>
<thead>
<tr>
<th>Yindjibarndi</th>
<th>Dictionary</th>
<th>“Correct”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wundu</td>
<td>River(bed)</td>
<td>Riverbed</td>
</tr>
<tr>
<td>Yijirdi</td>
<td>small stream, rivulet</td>
<td>Trickle of water</td>
</tr>
<tr>
<td>Manggurdu</td>
<td>Fortescue River</td>
<td>Strong, deep flow</td>
</tr>
</tbody>
</table>
Watercourse Summary

• The English ‘ontology’ has sizes of watercourses: rivers and creeks
• (Watercourses have beds)
• Yindjibarndi ‘ontology’ has two sizes of (dry) fluvial channels: wundu and gargā
• Yindjibarndi ‘ontology’ also has two ‘sizes’ of flow events, strong deep flow (manggurdu) and trickles (yijirdi)
Terms for water features

• In Yindjibarndi permanent and temporary water features that otherwise appear similar are considered to be different kinds of features.

• English, in contrast, treats permanence of water bodies and water courses as an attribute, and expresses it through adjectives like "temporary", "seasonal", "intermittent", or "ephemeral".

• Key distinction in English is *still* vs *flowing* water.

• It seems (from recent fieldwork) that permanent water in Yindjibarndi is a *yinda*, whether still or flowing. Thus there is a significant difference between conceptualizations of water features.
Some General Conclusions from Yindjibarndi Study

• None of the Yindjibarndi terms for landscape features that we have examined in depth so far is exactly equivalent to one single term in English
• Yindjibarndi terms divide up sub-domains of geographic reality (somewhat) differently than do English terms
• The results support the basis Ethnophysiography hypothesis:
  • people from different places and cultures apparently use different conceptual categories for geographic features
Study #4: Rating Landform Pictures Against Terms

- As part of a larger study of geospatial ontology, we asked subjects to rate how well a given landform term was exemplified by a particular picture.
- Each subject was given one term and six photographs.
Human Subjects Testing

• Each subject was given a page with six pictures of mountains, hills, etc., and one term that they were to compare to each photograph.
• Four different pages of photos with 6 photos each were used, and three or four different terms were tested.
The Subjects

- **English**: 334 subjects in Buffalo, with at least 25 judgments for each of the 16 photographs tested
- **Spanish**: 305 subjects in Guatemala City, with at least 13 judgments for each of the 16 photographs tested
- **German**: 65 subjects in Münster, with at least 20 judgments for each of the 6 photographs tested
The Photographs
The Terms

• **English**: Mountain, volcano, hill
• **Spanish**: Montaña, volcán, colina, cerro
• **German**: Berg, Vulkan, Hügel
The Ratings

• “For each photograph, please rate how good an example of a mountain it shows, on a scale from excellent (7) to poor (1)”

• Average scores were transformed so that unanimous “excellent” became 1.00 and unanimous “poor” became 0.00
Excellent Examples of Mountains
Good Examples of Mountains
Borderline Example of a Mountain
Poor Examples of Mountains
Very Poor Examples of Mountains
Volcano
(English)
Vulkan
(German)
Volcano

Spanish

German

English

[Diagram with numbers and images representing different languages and their corresponding volcano concepts]
Volcano

Spanish
German
English

0.94
0.91
0.92

0.52 0.44 0.31 0.17 0.12
0.33 0.28 0.19 0.00
0.58 0.39 0.36 0.10 0.03

1.00
0.50
0.00

Languages:
- Spanish
- German
- English
Volcano

Spanish

German

English

0.94

0.91

0.92

0.33

0.28

0.19

0.00

0.52 0.44 0.31 0.17 0.12

0.33

0.28

0.19

0.00

0.58 0.39 0.36 0.10 0.03

1.00

0.50

0.00
<table>
<thead>
<tr>
<th>Language</th>
<th>Volcano</th>
<th>English</th>
<th>German</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.94</td>
<td>0.52</td>
<td>0.33</td>
<td>0.33</td>
<td>0.94</td>
</tr>
<tr>
<td>0.91</td>
<td>0.44</td>
<td>0.28</td>
<td>0.28</td>
<td>0.91</td>
</tr>
<tr>
<td>0.92</td>
<td>0.31</td>
<td>0.19</td>
<td>0.19</td>
<td>0.92</td>
</tr>
<tr>
<td>0.17</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>0.36</td>
<td>0.52</td>
<td>0.10</td>
<td>0.10</td>
<td>0.36</td>
</tr>
<tr>
<td>0.03</td>
<td>0.31</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Volcán Agua, Guatemala
A “Hill” in England
Conclusions for Study #4

• For some images, the terms appear to translate well (i.e., similar ratings in all three languages)—robust cross-linguistic prototypes?
• For other images, ratings vary substantially across languages
• More research needed!
Summary and Conclusions

- Two kinds of research are needed before we can implement the Spatial Semantic Web for Landscape:
- One is to compile formal definitions of landscape feature types in languages of interest
- The other is to develop computational methods for delimiting and classifying landscape features from DEMs and images
A ‘Middle-ware’ of fundamental landform concepts

Concept 1

“mountain”

Concept 2

“montaña”

Concept 3

“mont”

Concept 4

“berg”

Concept 5

“山”

Real world

What is needed?

Geospatial information
• The landform concept modules (with weights and parameters) could be combined to provide formal definitions of the concepts associated with each landscape term.
A ‘Middle-ware’ of fundamental landform concepts

- The **same** landform concept modules can also be ‘executed’ against the DEM to detect and delimit regions that have the attributes embodied in the modules, and thus detect and delimit landform instances.
Thanks! (For more information, see)
http://www.geog.buffalo.edu/ncgia/ethnophysiography/
http://www.geog.buffalo.edu/ncgia/ontology/

This research was supported in part by NSF Grants BCS-9975557 (David Mark and Barry Smith) and BCS-0423075 and BCS-0423023 (David Mark and David Stea)