How Gazetteers Work: Cultural and Linguistic Influences on the Georeferencing Process

David M. Mark
NCGIA- Buffalo
Georeferencing as a Process

• Georeferencing as a *cognitive* process
• Georeferencing as a *computational* process
• Georeferencing for a human user (information retrieval)
  – General public
  – Experts, professionals
• Georeferencing as a service for other computer applications (semantic web, etc.)

• In what ways (if any!) should knowledge of the cognitive process inform the design of computational solutions?
Context?

• How different are the requirements for gazetteers in different contexts?
• Digital Libraries?
• Information retrieval? GeoCLEF etc.
• Single language or multilingual?
  – Sources? Users?
This project is being done in collaboration with
  - David Stea (Texas State, San Marcos)
  - Andrew Turk (Murdoch University, Perth, Australia)
  - Carmelita Topaha (San Juan College)
  - the Yindjibarndi people (Roebourne, Australia)
  - Navajo people and the Navajo Nation

The project is supported by National Science Foundation Grants BCS-0423075 (to David Mark) and BCS-0423023 (to David Stea), by the National Geospatial-intelligence Agency, and by Microsoft Research

The Navajo part is being conducted under permit C0513-E. Historic Preservation Office, Navajo Nation

The Ethnophysiology Project
The Ethnophysiography Project: Five Inter-related Topics

- **"Geographic Categories"**: Common nouns or noun-phrases that refer to *kinds* of geographic things ("ethnophysiography")
- **Toponyms**: Proper names for individual geographic features
- **Indigenous Geographic Knowledge Systems**: For example, traditional stories that incorporate landscape features (cf. Keith Basso)
- **Topophilia**: Emotional bonds between people, place, and landscape (cf. Yi-fu Tuan)
- **Indigenous Mapping, Indigenous GOS**
Pre-History of Gazetteers

• Before written language and graphic maps, geographic information was often stored and transmitted in *stories*, which often also included origin stories or moral codes.

• Place names, and the places themselves, formed ‘retrieval keys’ for the information.
Blue: Journey of Younger Sister, Wilito Wilson's version
Red: Journey of Younger Sister, Singer Man's version
“Wisdom Sits in Places”

- The role of landscape and place in the representation of traditional knowledge in oral traditions
The Gazetteer Triangle:
“Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints”
The Gazetteer Triangle:
“Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints”

- To know how gazetteers work, we need to understand each corner of this triangle
The Semiotic Triangle
("The Meaning Triangle")

Semiotic Triangle on the Instance

Concept of the Instance

The Actual Instance

Place Name
Semiotic Triangle on the Category

Concept of the Category

Geographic Category

Extension of the Category (Instances)
Three Triangles

Concept of the Category

Place Name

Footprint

Extension of the Category (Instances)

Geographic Category

The Actual Instance

Concept of the Instance

Footprint
Instance-to-class (Is-A) for Concept
Instance-to-class (Is-A) for Reality

Concept of the Instance

The Actual Instance

Extension of the Category (Instances)
Location of the Instance

The Actual Instance -> Footprint
Delimitation of the Instance

- Concept of the Category
  - Geographic Category
  - Delimitation details may depend on the Category (crisp or graded, etc.)
- The Actual Instance
- The Delimitation Problem
- Footprint
Cultural Differences in Place Names

The Actual Instance

Place Name

Footprint

Concept of the Category

Geographic Category

Extension of the Category (Instances)

Concept of the Instance
Cultural Differences in Categories

- Concept of the Category
- Geographic Category
- Extension of the Category
- Concept of the Instance
- The Actual Instance
- Place Name
- Footprint
Cultural Differences in Delimitation?

Concept of the Category

Geographic Category

Extension of the Category

Concept of the Instance

The Actual Instance

Place Name

Footprint

Delimitation
Aside:
Do we really need categories in a “Geo-referencing System”?
Cultural & Linguistic Differences

• My main research the last four years has been on cultural and linguistic differences in geospatial conceptualization and referencing
Proper Names

• I won’t say much about the “names” corner
• Except for this sobering complication: for the Navajo, many (most?) geographic features have at least two different proper names: one for traditional (origin) stories and one for every-day use
• It gets even more ‘interesting’: at least some of the sacred names are only to be used in the winter!
Categories: Concepts with no Word in English

Here are some English-language descriptions of some landscape features that have single words in some other language, but no single word in English:

- “A canyon wall receiving sunlight”
- “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”
- “An ‘island’ of land completely surrounded by one or more younger lava flows”
- “An island of grassland left unburnt after a surrounding wildfire”
Example: Hawai’ian has a word “Kipuka”

- Kipuka: A Hawai’ian word for an ‘island’ of land completely surrounded by one or more younger lava flows
• Icelandic Has a Word for a lava ‘island’ too!
  – Óbrinnishólmi
  – Literally, “un-burnt-hill”

• And the Walmajarri (in Australia) language has a word with a similar meaning:
  – Nyirrr: an island of grass left unburnt by surrounding fire
• So, if we based our ideas about geographic categories only on the meanings of the words in English, we would miss out on a lot!
Semantics and Categories

• “Show the ten largest lakes in France and provide a table of their statistics”
# Standing Water, English-French, Naïve view

<table>
<thead>
<tr>
<th></th>
<th>Lac</th>
<th>Étang</th>
<th>Lagune</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pond</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lagoon</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Semantics and Categories

• “Show the ten largest lakes in France and provide a table of their statistics” perhaps becomes something like…

• \(<\langle\text{France}\rangle<\text{lac}\rangle<\text{contained}\rangle<\text{area}\rangle>\)

• But… does “lac” == “lake”?
French and English Categorize Standing Water Bodies Differently
# Standing Water, English-French, More complete view

<table>
<thead>
<tr>
<th></th>
<th>Lac</th>
<th>Étang</th>
<th>Lagune</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Pond</td>
<td>?</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Lagoon</td>
<td>?</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Semantics and Categories, B

• “Show the ten largest rivers in France and provide a table of their statistics” perhaps becomes…

• $\langle\langle\text{France}\rangle\langle\text{fleuve}\rangle\langle\text{contained}\rangle\langle\text{length}\rangle\rangle$
# Liste des fleuves de France

Un article de Wikipédia, l'encyclopédie libre.

(Redirigé depuis Fleuve français)

Ces deux tableaux présentent le classement des fleuves français, selon leur longueur sur le territoire national.

## Sommaire

- 1 France métropolitaine
  - 1.1 Fleuves
  - 1.2 Fleuves côtiers
- 2 Guyane
- 3 Réunion
- 4 Voir aussi
  - 4.1 Notes et références

## France métropolitaine

### Fleuves

Liste des fleuves métropolitains, ordonnés par longueur (sur le territoire français) décroissante :

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur totale (km)</th>
<th>Longueur en France (km)</th>
<th>Adjectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Atlantique</td>
<td>1 012</td>
<td>1 012</td>
<td>ligérien(ne)</td>
</tr>
</tbody>
</table>
### Fleuves

Liste des fleuves métropolitains, ordonnés par longueur (sur le territoire français) décroissante :

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur totale (km)</th>
<th>Longueur en France (km)</th>
<th>Adjectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Atlantique</td>
<td>1 012</td>
<td>1 012</td>
<td>ligérien(ne)</td>
</tr>
<tr>
<td>Seine</td>
<td>Manche</td>
<td>776</td>
<td>776</td>
<td>séquanais(e)</td>
</tr>
<tr>
<td>Garonne</td>
<td>Atlantique</td>
<td>575</td>
<td>530</td>
<td>garonnais(e)</td>
</tr>
<tr>
<td>Rhône</td>
<td>Méditerranée</td>
<td>812</td>
<td>522</td>
<td>rhodanien(ne)</td>
</tr>
<tr>
<td>Meuse</td>
<td>Mer du Nord</td>
<td>950</td>
<td>483</td>
<td>mosan(e)</td>
</tr>
<tr>
<td>Charente</td>
<td>Atlantique</td>
<td>360</td>
<td>360</td>
<td>charentais(e)</td>
</tr>
<tr>
<td>Adour</td>
<td>Atlantique</td>
<td>335</td>
<td>335</td>
<td>adourien(ne)</td>
</tr>
<tr>
<td>Somme</td>
<td>Manche</td>
<td>245</td>
<td>245</td>
<td>?</td>
</tr>
<tr>
<td>Vilaine</td>
<td>Atlantique</td>
<td>225</td>
<td>225</td>
<td>?</td>
</tr>
<tr>
<td>Aude</td>
<td>Méditerranée</td>
<td>220</td>
<td>220</td>
<td>?</td>
</tr>
<tr>
<td>Rhin</td>
<td>Mer du Nord</td>
<td>1 325</td>
<td>195</td>
<td>rhénan(e)</td>
</tr>
<tr>
<td>Escaut</td>
<td>Mer du Nord</td>
<td>430</td>
<td>180</td>
<td>scaldien(ne)</td>
</tr>
</tbody>
</table>

### Fleuves côtiers

Liste des principaux fleuves côtiers métropolitains, ordonnés par longueur décroissante :

http://fr.wikipedia.org/wiki/Fleuve_fran%C3%A7ais
What about the Marne and the Dordogne??

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur totale (km)</th>
<th>Longueur en France (km)</th>
<th>adjectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Atlantique</td>
<td>1 012</td>
<td>1 012</td>
<td>ligérien(ne)</td>
</tr>
<tr>
<td>Seine</td>
<td>Manche</td>
<td>776</td>
<td>776</td>
<td>séquanais(e)</td>
</tr>
<tr>
<td>Garonne</td>
<td>Atlantique</td>
<td>575</td>
<td>530</td>
<td>garonnais(e)</td>
</tr>
<tr>
<td>Rhône</td>
<td>Méditerranée</td>
<td>812</td>
<td>522</td>
<td>rhodanien(ne)</td>
</tr>
<tr>
<td>Meuse</td>
<td>Mer du Nord</td>
<td>950</td>
<td>483</td>
<td>mosan(e)</td>
</tr>
<tr>
<td>Charente</td>
<td>Atlantique</td>
<td>360</td>
<td>360</td>
<td>charentais(e)</td>
</tr>
<tr>
<td>Adour</td>
<td>Atlantique</td>
<td>335</td>
<td>335</td>
<td>adourien(ne)</td>
</tr>
<tr>
<td>Somme</td>
<td>Manche</td>
<td>245</td>
<td>245</td>
<td>?</td>
</tr>
<tr>
<td>Vilaine</td>
<td>Atlantique</td>
<td>225</td>
<td>225</td>
<td>?</td>
</tr>
<tr>
<td>Aude</td>
<td>Méditerranée</td>
<td>220</td>
<td>220</td>
<td>?</td>
</tr>
<tr>
<td>Rhin</td>
<td>Mer du Nord</td>
<td>1 325</td>
<td>195</td>
<td>rhénan(e)</td>
</tr>
<tr>
<td>Escaut</td>
<td>Mer du Nord</td>
<td>430</td>
<td>180</td>
<td>scaldien(ne)</td>
</tr>
</tbody>
</table>

Marne: 525 km
Dordogne: 490 km
The Marne and the Dordogne are "rivières", not "fleuves"!

Obvious implications for multilingual information retrieval à la GeoCLEF!
Recently, we have been using two main methods in our fieldwork:

• **“Field interviews”:** collecting language actually used while out in the landscape (10 interviews so far; 28 hours of audio!)
• **“Photo Response”:** Showing landscape photographs to people and getting them to talk about them (7 sessions, 9 hours of audio in Navajo, 5 hours in Yindjibarndi)
Loretta Holyan, Andrew Turk, Carmelita Topaha, Larry King

September 25, 2005
Blue Canyon, Arizona
Amelie Collins, Erma Batteso, Louise Hubbell, Harding Yazzie Jr.  
March 2006, San Juan College, Farmington
• Jin-kyu Jung has linked photos and metadata to Google Maps
• Aiming for links to transcripts, audio
• Linking to ATLAS.ti for qualitative analysis
Footprints

- There has been very little ‘human subjects’ work on delimitation of geographic features
Right now, we are in one of the exceptions!

Individual Responses

- 100% confidence regions

- 50% confidence regions
Combined, Dot-density

- 100% confidence regions
- 50% confidence regions
Landforms

• Claim: A landform is a shape-based part of the Earth's surface, occupying a finite region, that has some degree of perceptual or functional coherence of form (shape)
• This definition has some circularity:
  – an arbitrary region is unlikely to contain topography that has a coherent shape
  – yet without specifying the region in advance, there is no ‘object’ for which shape can be determined
Ontology of ‘Landform’ in General

• In the case of landforms, the shape of the landform is inherited from the pattern of elevations of the Earth's crust within its boundaries.

• Thus, land surface shape and landform boundary are mutually dependent.
(A Note on the Goals of the Project)

- The goal is to understand the meanings of words used by ‘general’ speakers of languages to refer to landforms, words such as hill, hügel, colina, cerro, loma, berg, mountain, butte, mesa, marnda, etc.
- We are not especially interested (at this time) in scientific vocabularies for landform types that might be used in Earth science
A Complication

• Cross-cultural comparisons of landform definitions have found cultural and linguistic differences in landform types (Mark and Turk 2003)

• If the basic categories and even their relationships differ across languages, different taxonomic hierarchies may be needed for the same region of the Earth’s surface, one for each language or dialect or culture
Multilingual Ontology

• If a goal of the geographic aspects of the Semantic Web and the internet is to provide on-line access to geographic information about any area of the world to speakers of any language, we might have to record feature codes (entity types) separately for each language!

• However, a better approach would be to develop procedures for feature delimitation and classification that could accept a formalized version of any landform definition, and then extract and classify instances from digital elevation data
Topographic Eminences

• In this paper, we present first steps toward such a system by outlining a strategy for definition-dependent extraction of topographic eminences

• We have chosen the term “eminence” as a super-ordinate term for hills, mountains, buttes, mesas, and other such landforms that stand above their immediate neighborhoods, because no other English word fits, and because Voegelin and Voegelin (“Hopi Domains”) used eminence in this sense
Topographic Eminences

• Perceptual characteristics such as size, shape, color, texture and material character may influence the perception and classification of eminences
• Whereas shape and size alone may not always be sufficient to define eminence types completely, we believe that size and shape are the most important factors that determine how people perceive and categorize eminences
• Today, we focus on form-based extraction and classification of eminences for categories commonly recognized in ‘every-day’ English
Hills and Mountains

• Mountain:

1. a. A large natural elevation of the earth's surface, *esp.* one high and steep in form (larger and higher than a hill) (OED)
Hills and Mountains

- **Hill:**
  1. a. A natural elevation of the earth's surface … after the introduction of the word mountain’ [into English], gradually restricted to heights of less elevation; … (OED)
Hills and Mountains: Not Only Size!

• Hill: “a more rounded and less rugged outline is also usually connoted by the name” (OED)

Hill  ??  Mountain
For features too small to be ‘mountains’, yet too jagged to be ‘hills’, English relies on other terms, such as rock, butte, peak, mesa, etc.
Extraction of Eminences

- Wood (1996) identified morphometric peaks as representatives of eminences and later defined summit regions as inclusive of all points on the eminence within a user-specified elevation difference from the peaks (Wood 2004)
- Summit regions can also be defined and delimited by *prominence*, defined as the elevation difference from the summit to the highest neighboring saddle
Extraction of Eminences

- Summit-driven eminence detection defines the core of the eminence.
- ‘Uphill’ catchments provide one approach to delineate the lower boundaries of eminences; these are exactly hills as defined in 1870 by James Clerk Maxwell (‘On Hills and Dales’).
- Slope, curvature, and breaks of slope might also be used to locate the outer boundaries of eminences.
Extraction of Eminences: An Example

- Mount Washington area, Presidents Range, New Hampshire
- Part of Gaurav Sinha’s dissertation
Classification

• Once eminences are extracted, they can be classified based on properties such as:
  – *i*) *Morphographic character*—shape, size, position, orientation
  – *ii*) *Spatial relationships*—proximity, prominence, topology
  – *iii*) *Visual signature*—viewshed statistics, angle of depression/elevation, visual prominence.
Classification

• A combination of these characteristics can be determined for eminences extracted by the methods sketched above, and a parameter space can be used to determine whether eminences vary continuously, or whether they cluster; each cluster is a candidate for a subclass of eminence

• A lack of clear clusters would suggest that landforms form continua and do not form 'natural kinds', but that landform categories are based on perceptual and utilitarian (affordance-based) factors
Future Work on Feature Extraction

• Determine and implement a suite of measures of the 3-dimensional shapes of eminences
• Look for clusters of features in a parameter space
• Try to distinguish eminence types named and used in English, based on these measures
• Apply the method to DEMs from the Navajo reservation
• See whether eminence types named and used in the Navajo can also be defined from there parameters of detected in the parameter space
• Extend to other regions and other languages
• Eventually, develop methods for other higher-level landform types such as canyons and valleys
The Challenge

• There are about 5,000 languages in the world that still have 1,000 or more speakers
• There may be about 100 ‘geographical’ terms per language
• 500,000 terms that need to be defined and implemented!
Thanks!

For more information, see http://www.ncgia.buffalo.edu/ethnophysiography/
How Gazetteers Work:
Cultural and Linguistic Influences on the Georeferencing Process

David M. Mark
NCGIA- Buffalo
Georeferencing as a Process

- Georeferencing as a **cognitive** process
- Georeferencing as a **computational** process
- Georeferencing for a human user (information retrieval)
  - General public
  - Experts, professionals
- Georeferencing as a service for other computer applications (semantic web, etc.)

- In what ways (if any!) should knowledge of the cognitive process inform the design of computational solutions?
Context?

• How different are the requirements for gazetteers in different contexts?
• Digital Libraries?
• Information retrieval? GeoCLEF etc.
• Single language or multilingual?
  – Sources? Users?
• This project is being done in collaboration with
  – David Stea (Texas State, San Marcos)
  – Andrew Turk (Murdoch University, Perth, Australia)
  – Carmelita Topaha (San Juan College)
  – the Yindjibarndi people (Roebourne, Australia)
  – Navajo people and the Navajo Nation

The project is supported by
National Science Foundation Grants BCS-0423075 (to David Mark) and BCS-0423023 (to David Stea), by the National Geospatial-intelligence Agency, and by Microsoft Research

The Navajo part is being conducted under permit C0513-E, Historic Preservation Office, Navajo Nation
The Ethnophysiography Project: Five Inter-related Topics

- **“Geographic Categories”**: Common nouns or noun-phrases that refer to kinds of geographic things (“ethnophysiography”)
- **Toponyms**: Proper names for individual geographic features
- **Indigenous Geographic Knowledge Systems**: For example, traditional stories that incorporate landscape features (cf. Keith Basso)
- **Topophilia**: Emotional bonds between people, place, and landscape (cf. Yi-fu Tuan)
- **Indigenous Mapping, Indigenous GIS**
Pre-History of Gazetteers

• Before written language and graphic maps, geographic information was often stored and transmitted in stories, which often also included origin stories or moral codes.

• Place names, and the places themselves, formed ‘retrieval keys’ for the information.
“Wisdom Sits in Places”

- The role of landscape and place in the representation of traditional knowledge in oral traditions
The Gazetteer Triangle:
“Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints”
The Gazetteer Triangle:
“Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints”

- To know how gazetteers work, we need to understand each corner of this triangle.
The Semiotic Triangle
(“The Meaning Triangle”)

Semiotic Triangle on the Instance

Concept of the Instance

The Actual Instance

Place Name
Semiotic Triangle on the Category

Concept of the Category

Geographic Category

Extension of the Category (Instances)
Three Triangles

Geographic Category

Concept of the Category (Instances)

Extension of the Category

Concept of the Instance

The Actual Instance

Place Name

Footprint
Instance-to-class (Is-A) for Concept
Instance-to-class (Is-A) for Reality
Location of the Instance

The Actual Instance

Footprint
The Delimitation of the Instance

Geographic Category

Concept of the Category

Delimitation details may depend on the Category (crisp or graded, etc.)

The Actual Instance

The Delimitation Problem

Footprint
Cultural Differences in Place Names

- Concept of the Category
- Geographic Category
- Extension of the Category (Instances)
- The Actual Instance
- Place Name
- Footprint

Extension of the Category (Instances)
Cultural Differences in Categories

Concept of the Category

Geographic Category

Extension of the Category

The Actual Instance

Place Name

Footprint
Cultural Differences in Delimitation?

- Concept of the Category
- Geographic Category
- Extension of the Category

Concept of the Instance

The Actual Instance

Place Name

Footprint

Delimitation
Aside:
Do we really need categories in a “Geo-referencing System”?

Concept of the Category

Geographic Category

Extension of the Category

Concept of the Instance

The Actual Instance

Place Name

Footprint
Cultural & Linguistic Differences

• My main research the last four years has been on cultural and linguistic differences in geospatial conceptualization and referencing
Proper Names

- I won’t say much about the “names” corner
- Except for this sobering complication: for the Navajo, many (most?) geographic features have at least two different proper names: one for traditional (origin) stories and one for every-day use
- It gets even more ‘interesting’: at least some of the sacred names are only to be used in the winter!
Categories:
Concepts with no Word in English

• Here are some English-language descriptions of some landscape features that have single words in some other language, but no single word in English
  – “A canyon wall receiving sunlight”
  – “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”
  – “An ‘island’ of land completely surrounded by one or more younger lava flows”
  – “An island of grassland left unburnt after a surrounding wildfire”
Example: Hawai‘ian has a word “Kipuka”

- Kipuka: A Hawai‘ian word for an ‘island’ of land completely surrounded by one or more younger lava flows
• Icelandic Has a Word for a lava 'island' too!
  – Óbrinnishólmi
  – Literally, "un-burnt-hill"

• And the Walmajarri (in Australia) language has a word with a similar meaning:
  – Nylirim: an island of grass left unburnt by surrounding fire
So, if we based our ideas about geographic categories only on the meanings of the words in English, we would miss out on a lot!
Semantics and Categories

• “Show the ten largest lakes in France and provide a table of their statistics”
### Standing Water, English-French, Naïve view

<table>
<thead>
<tr>
<th></th>
<th>Lac</th>
<th>Étang</th>
<th>Lagune</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pond</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lagoon</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Semantics and Categories

• “Show the ten largest lakes in France and provide a table of their statistics” perhaps becomes something like…

• <<France><lac><contained><area>>

• But… does “lac” == “lake”?
French and English Categorize Standing Water Bodies Differently
### Standing Water, English-French, More complete view

<table>
<thead>
<tr>
<th></th>
<th>Lac</th>
<th>Étang</th>
<th>Lagune</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Pond</td>
<td>?</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Lagoon</td>
<td>?</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Semantics and Categories, B

• “Show the ten largest rivers in France and provide a table of their statistics" perhaps becomes…

• $\langle\langle$France$\rangle<$fleuve$<$contained$><$length$>$>
Liste des fleuves de France

Un article de Wikipédia, l'encyclopédie libre.

(Relevé depuis l'encyclopédie libre)

Ce document présente le classement des fleuves français, selon leur longueur sur leur territoire national.

Sommaire

- France métropolitaine
  - 1 Fleuves
  - 2 Fleuves en France
  - 3 Fleuves
  - 4 Notes et références

France métropolitaine

Fleuves

Liste des fleuves métropolitains, ordonnés par longueur sur le territoire français, suivants :

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur en France (km)</th>
<th>Longueur sur l'état correspondant</th>
<th>État correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Rhône</td>
<td>1 012</td>
<td>1 012</td>
<td>France</td>
</tr>
<tr>
<td>L'Isère</td>
<td>Rhône</td>
<td>1 012</td>
<td>1 012</td>
<td>France</td>
</tr>
</tbody>
</table>
### Fleuves

Liste des fleuves métropolitains, ordonnés par longueur (sous forme de km):

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur total (km)</th>
<th>Longueur vers l'Atlantique (km)</th>
<th>Né au(e)</th>
<th>Objectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Atlantique</td>
<td>1 012</td>
<td>934</td>
<td>638</td>
<td>Loire</td>
</tr>
<tr>
<td>Saône</td>
<td>Rhône</td>
<td>776</td>
<td>716</td>
<td>660</td>
<td>Rhône</td>
</tr>
<tr>
<td>Garonne</td>
<td>Méditerranée</td>
<td>575</td>
<td>500</td>
<td>430</td>
<td>Garonne</td>
</tr>
<tr>
<td>Rhône</td>
<td>Méditerranée</td>
<td>812</td>
<td>632</td>
<td>522</td>
<td>Rhône</td>
</tr>
<tr>
<td>Meuse</td>
<td>Rhône</td>
<td>950</td>
<td>838</td>
<td>720</td>
<td>Meuse</td>
</tr>
<tr>
<td>Charente</td>
<td>Atlantique</td>
<td>360</td>
<td>300</td>
<td>240</td>
<td>Charente</td>
</tr>
<tr>
<td>Adour</td>
<td>Atlantique</td>
<td>335</td>
<td>295</td>
<td>215</td>
<td>Adour</td>
</tr>
<tr>
<td>Seine</td>
<td>Meuse</td>
<td>245</td>
<td>215</td>
<td>195</td>
<td>Seine</td>
</tr>
<tr>
<td>Vienne</td>
<td>Atlantique</td>
<td>220</td>
<td>220</td>
<td>190</td>
<td>Vienne</td>
</tr>
<tr>
<td>Ardèche</td>
<td>Méditerranée</td>
<td>220</td>
<td>220</td>
<td>190</td>
<td>Ardèche</td>
</tr>
<tr>
<td>Rhône</td>
<td>Rhône</td>
<td>1 305</td>
<td>1 095</td>
<td>995</td>
<td>Rhône</td>
</tr>
<tr>
<td>Eure</td>
<td>Rhône</td>
<td>430</td>
<td>430</td>
<td>380</td>
<td>Eure</td>
</tr>
</tbody>
</table>

### Fleuves côtiers

Liste des principaux fleuves côtiers métropolitains, ordonnés par longueur décroissante:

https://fr.wikipedia.org/wiki/Liste_des_fleuves_c%C3%AAtiers_m%C3%A9tropolitains

Page 1 of 1
**Fleuves**

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur total (km)</th>
<th>Longueur en France (km)</th>
<th>Objectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire</td>
<td>Atlantique</td>
<td>1 512</td>
<td>1 012</td>
<td>Loire</td>
</tr>
<tr>
<td>Seine</td>
<td>Mer hermine</td>
<td>770</td>
<td>770</td>
<td>Mer hermine</td>
</tr>
<tr>
<td>Garonne</td>
<td>Atlantique</td>
<td>570</td>
<td>520</td>
<td>Garonne</td>
</tr>
<tr>
<td>Rhône</td>
<td>Méditerranée</td>
<td>812</td>
<td>520</td>
<td>Rhône</td>
</tr>
<tr>
<td>Meuse</td>
<td>Ruisseau de la Seine</td>
<td>950</td>
<td>483</td>
<td>Meuse</td>
</tr>
<tr>
<td>Cherize</td>
<td>Atlantique</td>
<td>360</td>
<td>300</td>
<td>Cherize</td>
</tr>
<tr>
<td>Adour</td>
<td>Atlantique</td>
<td>335</td>
<td>315</td>
<td>Adour</td>
</tr>
<tr>
<td>Soustes</td>
<td>Mer hermine</td>
<td>245</td>
<td>245</td>
<td>Soustes</td>
</tr>
<tr>
<td>Vivette</td>
<td>Atlantique</td>
<td>205</td>
<td>205</td>
<td>Vivette</td>
</tr>
<tr>
<td>Audi</td>
<td>Méditerranée</td>
<td>200</td>
<td>200</td>
<td>Audi</td>
</tr>
<tr>
<td>Rhône</td>
<td>Ruisseau de la Seine</td>
<td>1 325</td>
<td>195</td>
<td>Rhône</td>
</tr>
<tr>
<td>Eure</td>
<td>Ruisseau de la Seine</td>
<td>430</td>
<td>195</td>
<td>Eure</td>
</tr>
</tbody>
</table>

**Fleuves côtiers**

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Tributaire de</th>
<th>Longueur total (km)</th>
<th>Longueur en France (km)</th>
<th>Objectif correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loire côtiers</td>
<td>Atlantique</td>
<td>1 512</td>
<td>1 012</td>
<td>Loire côtiers</td>
</tr>
<tr>
<td>Seine côtiers</td>
<td>Mer hermine</td>
<td>770</td>
<td>770</td>
<td>Mer hermine côtiers</td>
</tr>
<tr>
<td>Garonne côtiers</td>
<td>Atlantique</td>
<td>570</td>
<td>520</td>
<td>Garonne côtiers</td>
</tr>
<tr>
<td>Rhône côtiers</td>
<td>Méditerranée</td>
<td>812</td>
<td>520</td>
<td>Rhône côtiers</td>
</tr>
</tbody>
</table>

*What about the Marne and the Dordogne??*

Marne: 525 km
Dordogne: 490 km
The Marne and the Dordogne are "rivières", not "fleuves"!

<table>
<thead>
<tr>
<th>Cours d'eau</th>
<th>Affluent de</th>
<th>Longueur</th>
<th>Bassin versant</th>
<th>Objet correspondant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meuse</td>
<td>Saône</td>
<td>520 km</td>
<td>12 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Dordogne</td>
<td>Garonne</td>
<td>480 km</td>
<td>24 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Lot</td>
<td>Garonne</td>
<td>481 km</td>
<td>17 440 km²</td>
<td>-</td>
</tr>
<tr>
<td>Loire</td>
<td>Rance</td>
<td>480 km</td>
<td>29 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Dordogne</td>
<td>Saône</td>
<td>480 km</td>
<td>7 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Yonne</td>
<td>Loire</td>
<td>410 km</td>
<td>15 300 km²</td>
<td>-</td>
</tr>
<tr>
<td>Tarn</td>
<td>Garonne</td>
<td>390 km</td>
<td>15 700 km²</td>
<td>-</td>
</tr>
<tr>
<td>Vienne</td>
<td>Loire</td>
<td>370 km</td>
<td>20 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Cher</td>
<td>Loire</td>
<td>220 km</td>
<td>14 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Orthez</td>
<td>Rance</td>
<td>300 km</td>
<td>14 225 km²</td>
<td>-</td>
</tr>
<tr>
<td>Moselle</td>
<td>Rhône</td>
<td>350 km</td>
<td>30 000 km²</td>
<td>-</td>
</tr>
<tr>
<td>Saône</td>
<td>Saône</td>
<td>315 km</td>
<td>12 005 km²</td>
<td>-</td>
</tr>
</tbody>
</table>

Obvious implications for multilingual information retrieval à la GeOCLEF!
Recently, we have been using two main methods in our fieldwork:

- **Field interviews**: collecting language actually used while out in the landscape (10 interviews so far; 28 hours of audio!)
- **Photo Response**: Showing landscape photographs to people and getting them to talk about them (7 sessions, 9 hours of audio in Navajo, 5 hours in Yindjabarndi)
Loretta Holyan, Andrew Turk, Carmella Topaha, Larry King
September 25, 2005
Blue Canyon, Arizona
Amelie Collins, Erma Batteo, Louise Hubbell, Harding Yazzie Jr.
March 2006, San Juan College, Farmington
• Jin-kyu Jung has linked photos and metadata to Google Maps
• Aiming for links to transcripts, audio
• Linking to ATLAS.ti for qualitative analysis
Footprints

• There has been very little ‘human subjects’ work on delimitation of geographic features
Right now, we are in one of the exceptions!

Individual Responses

- 100% confidence regions
- 50% confidence regions
Combined, Dot-density

- 100% confidence regions
- 50% confidence regions
• Claim: A landform is a shape-based part of the Earth’s surface, occupying a finite region, that has some degree of perceptual or functional coherence of form (shape)

• This definition has some circularity:
  – an arbitrary region is unlikely to contain topography that has a coherent shape
  – yet without specifying the region in advance, there is no ‘object’ for which shape can be determined
Ontology of ‘Landform’ in General

- In the case of landforms, the shape of the landform is inherited from the pattern of elevations of the Earth's crust within its boundaries.
- Thus, land surface shape and landform boundary are mutually dependent.
(A Note on the Goals of the Project)

• The goal is to understand the meanings of words used by ‘general’ speakers of languages to refer to landforms, words such as hill, hügel, colina, cerro, loma, berg, mountain, butte, mesa, marnda, etc.

• We are not especially interested (at this time) in scientific vocabularies for landform types that might be used in Earth science
A Complication

• Cross-cultural comparisons of landform definitions have found cultural and linguistic differences in landform types (Mark and Turk 2003)

• If the basic categories and even their relationships differ across languages, different taxonomic hierarchies may be needed for the same region of the Earth’s surface, one for each language or dialect or culture
Multilingual Ontology

• If a goal of the geographic aspects of the Semantic Web and the internet is to provide on-line access to geographic information about any area of the world to speakers of any language, we might have to record feature codes (entity types) separately for each language!

• However, a better approach would be to develop procedures for feature delimitation and classification that could accept a formalized version of any landform definition, and then extract and classify instances from digital elevation data.
Topographic Eminences

• In this paper, we present first steps toward such a system by outlining a strategy for definition-dependent extraction of topographic eminences

• We have chosen the term “eminence” as a super-ordinate term for hills, mountains, buttes, mesas, and other such landforms that stand above their immediate neighborhoods, because no other English word fits, and because Voegelin and Voegelin (“Hopi Domains”) used eminence in this sense
Topographic Eminences

- Perceptual characteristics such as size, shape, color, texture and material character may influence the perception and classification of eminences.
- Whereas shape and size alone may not always be sufficient to define eminence types completely, we believe that size and shape are the most important factors that determine how people perceive and categorize eminences.
- Today, we focus on form-based extraction and classification of eminences for categories commonly recognized in ‘every-day’ English.
Hills and Mountains

• **Mountain:**

  1. a. A large natural elevation of the earth's surface, *esp.* one high and steep in form (larger and higher than a hill) (OED)
Hills and Mountains

- **Hill:**
  1. a. A natural elevation of the earth's surface … after the introduction of the word mountain” [into English], gradually restricted to heights of less elevation; … (OED)
Hills and Mountains: Not Only Size!

- Hill: “a more rounded and less rugged outline is also usually connoted by the name” (OED)
For features too small to be ‘mountains’, yet too jagged to be ‘hills’, English relies on other terms, such as rock, butte, peak, mesa, etc.
Extraction of Eminences

- Wood (1996) identified morphometric peaks as representatives of eminences and later defined summit regions as inclusive of all points on the eminence within a user-specified elevation difference from the peaks (Wood 2004)
- Summit regions can also be defined and delimited by prominence, defined as the elevation difference from the summit to the highest neighboring saddle
Extraction of Eminences

• Summit-driven eminence detection defines the core of the eminence
• ‘Uphill’ catchments provide one approach to delineate the lower boundaries of eminences; these are exactly hills as defined in 1870 by James Clerk Maxwell (‘On Hills and Dales’)
• Slope, curvature, and breaks of slope might also be used to locate the outer boundaries of eminences
Extraction of Eminences: An Example

- Mount Washington area, Presidents Range, New Hampshire
- Part of Gaurav Sinha’s dissertation
Classification

• Once eminences are extracted, they can be classified based on properties such as:
  – i) **Morphographic character**—shape, size, position, orientation
  – ii) **Spatial relationships**—proximity, prominence, topology
  – iii) **Visual signature**—viewshed statistics, angle of depression/elevation, visual prominence.
Classification

• A combination of these characteristics can be determined for eminences extracted by the methods sketched above, and a parameter space can be used to determine whether eminences vary continuously, or whether they cluster; each cluster is a candidate for a subclass of eminence

• A lack of clear clusters would suggest that landforms form continua and do not form 'natural kinds', but that landform categories are based on perceptual and utilitarian (affordance-based) factors
Future Work on Feature Extraction

- Determine and implement a suite of measures of the 3-dimensional shapes of eminences
- Look for clusters of features in a parameter space
- Try to distinguish eminence types named and used in English, based on these measures
- Apply the method to DEMs from the Navajo reservation
- See whether eminence types named and used in the Navajo can also be defined from there parameters of detected in the parameter space
- Extend to other regions and other languages
- Eventually, develop methods for other higher-level landform types such as canyons and valleys
The Challenge

• There are about 5,000 languages in the world that still have 1,000 or more speakers
• There may be about 100 ‘geographical’ terms per language
• 500,000 terms that need to be defined and implemented!
Thanks!

For more information, see
http://www.ncgia.buffalo.edu/ethnophysiography/