

Topographic Subtyping of Place Named Entities : a linguistic approach

Van Tien Nguyen* Mauro Gaio* **Ludovic Moncla***

*Laboratoire LIUPPA, BP-1155, 64013 PAU Université Cedex
surname.name@univ-pau.fr

17 juin 2013

Outline

- 1 Problems and background
 - Movement or perception expressions
 - Named entities and spatial relations
- 2 Method and implementation
 - VT model proposal
 - Processing sequence
- 3 Experimentations and evaluation
- 4 Conclusion
 - Conclusion
 - Works in progress

Introduction

- Our works deals with Named Entity Recognition (NER) and Natural Language.
 - semantics and spatial representation of named places entities
- We use french travelogues, but we built methods and process based on cognitive principles.
- Our method can be adapted to differents languages by translating rules and patterns.

Introduction

Many systems of NER have been developed.
None of them are able to classify places into specific subtypes
such as RIVER, GLACIER, PEAK, MOUNTAIN.

Some NER tools

Dedicated to several types of named entities

- ANNIE, OpenCalais, Stanford NER, OpenNLP

Dedicated to named entities of type places

- MetaCarta, Yahoo ! Placemark

Dedicated to named entities of type dates

- GuTime, HeideTime

Problems and background

Different semantics and different spatial representation with the same named place entity.

Example : Artouste

- ... Artouste lake ...
- ... the peak of Artouste ...

Our goal

Find an existing sub-type (in the text) to reduce this ambiguity

Problems and background

Examples

Nous songeâmes bientôt à descendre sur le territoire aride de l' Aragon.

- Soon we were preparing to come down on the arid territory of Aragon

Nous arrivons au fond de la vallée d' Ossau.

- We arrived at the bottom of Ossau valley

- verbs of movement
- indirections (i.e. spatial relations)
- topographical terms
- place names

Problems and background

First contribution

Lexico-syntactic patterns to mark nominal groups having a topographical denotation.

- recognizes place named entities
- determines the semantics of the nominal group

Movement or perception expressions

In travelogues corpus, expression of movement is essential.

Verbs of movement : [Boo87], [Lau91], [Sar00]

- (1) initial verbs : to leave
- (2) final verbs : to arrive
- (3) median verbs : to cross

Verb of perception

- same behavior as median verbs : to see

(1)



(2)



(3)



Named entities and spatial relations

Linguistic models : [Tal00], [Van86] and [Bor98]

- concrete entity : denoted by topographical term
- spatial references : derived from place names

Spatial relation i.e. Indirections : [EF91], [Lig98]

- Au sud de, au nord de (south of, north of)
- A 10 km de (10 km from)
- a côté de, proche de (next to, near)

VT model proposal

Mark and formalise the relation between

- V : set of verbs (movement or perception)
- I : set of spatial relation (indirections)
- T : set of topographical terms
- G : set of place names

$VT = (v,t)$ with $v \in V$, $t \in (te, i, nt|t)$
where $te \in T$, $I \in I$ and $nt \in G$

Soon we were preparing to **come down** on the arid territory of Aragon

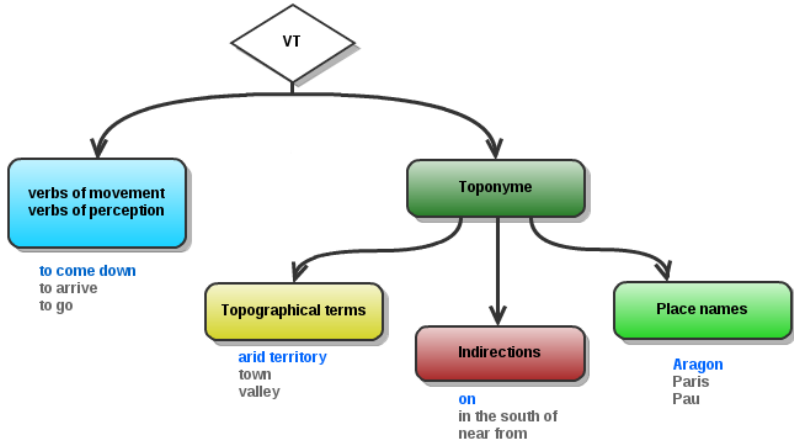


FIGURE : The VT structure

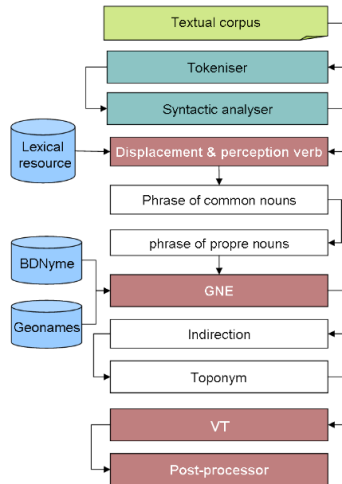
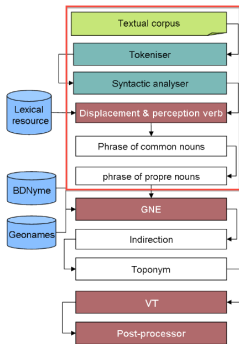


FIGURE : Processing main steps

Processing sequence



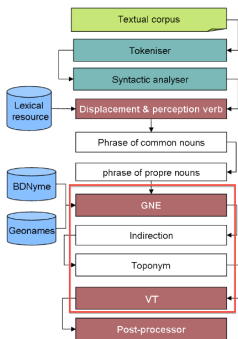
We made rules and lexico-syntactic patterns based on context-free-grammar (CFG) to :

- mark words or groups of words as common nouns or proper nouns
- replace a text sequence (noun, adjective, verb, etc), with a unique symbole

Example

term(adjectif : A .. noun : N) → adjectit(A), noun(N)

Mark the VT structure



- Gazetteers (BDNyme, Geonames) are used to validate proper nouns as existing Place Names
- Lexical resources are used to mark indirections

Toponyms are defined as a composition of the elements marked in previous steps.

Finally the VT structure is marked

$VT(\text{verb} : V .. \text{toponym} : T) \rightarrow \text{verb}(V), \text{toponyme}(T)$

Mark the VT structure

Examples

Nous songeâmes bientôt à descendre sur le territoire aride de l'Aragon.

- Soon we were preparing to come down on the arid territory of Aragon

Nous arrivons au fond de la vallée d'Ossau.

- We arrived at the bottom of Ossau valley

- VT structure
- verbs of movement
- indirections (i.e. spatial relations)
- common nouns
- place names

Determine the subtype for the place name

We use domain-specific ontology to find the subtypes.

Example

- We arrived at the bottom of Ossau **valley**.
 - I've crossed the largest **karstic** territory of Aragon.
- **"valley"** is part of the ontology so it becomes the subtype of the place name "Ossau".
 - **"karstic"** is not found on the ontology so it will not be directly used to sub-type the place name but it can be used to enrich the ontology.

Some experimentations

Our corpus : 14 French books (2400 pages)

- 105555 occurrences of verbs of movement
 - 1390 are involved in a VT pattern
- 560 VT patterns containing candidates for sub-typing Place Name.
 - 44 already exist in domain-specific ontology
 - 49 matched with key-concept in the RAMEAU thesaurus

Verbs of perception reveal new geographical information.

- 214 terms are connected to verbs of movement
- 68 terms are connected to verbs of perception
- 30% of terms only appear with verbs of perception

Conclusion

- We present a global method for adding sub-types to place named entities.
- This method can be used to
 - reduce ambiguities
 - improve queries for searching in very large resources
 - extract a lexicon of topographic labels

In travelogue corpus

The nominal group which lies between a verb (displacement or perception) and a place name, has a high probability of having a topographical sense.

Works in progress

- Improve the model and the VT structure (adding rules)
- Implement with other techniques (finite-state transducer)
- Make available for English and Spanish

Next step





Make a spatial representation of the movement not only of the named place entity

- Improve the model and the VT structure
- Find rules to find the origin and the destination in text





Questions I

Thank you very much !

Bibliographie I

-  J.-P Boons, *La notion sémantique de déplacement dans une classification syntaxique des verbes locatifs*, Langue Française (1987), no. 76, 5–40.
-  A. Borillo, *L'espace et son expression en français*.
-  Max J Egenhofer and Robert D Franzosa, *Point-set topological spatial relations*, International Journal of Geographical Information System **5** (1991), no. 2, 161–174.
-  D. Laur, *Sémantique du déplacement et de la localisation en français : une étude des verbes, des prépositions et de leurs relations dans la phrase simple*, Ph.D. thesis, 1991.

Bibliographie II

-  L. Ligozat, *Reasoning about cardinal directions*, J. Vis. Lang. Comput. **9** (1998), no. 1, 23–44.
-  L. Sarda, *L'expression du déplacement dans la construction transitive directe*, Syntaxe et sémantique (2000), no. 76, 121–137.
-  L. Talmy, *Toward a cognitive semantics*, edn, Mitt Press, chapter How language structures space, 2000.
-  C. Vandeloise, *edn, seuil, paris, frances*.