

Recherche des qualifiants dans le corpus de récits de voyage – Amélioration de la chaîne de traitement (Lot 1)

M.N. Bessagnet, M. Gaio, E. Kergosien,
T. Nguyen, C.Sallaberry

Equipe T2i
Laboratoire LIUPPA, UPPA



Contexte et Problématique

- Contexte
 - Recherche et extraction automatique de SN « qualifiants » des EN
 - Restriction à des EN de type lieu et à des qualifiants faisant référence à un concept géographique
- Problématique
 - Définir « les limites » d'une EN de type lieu
 - Catégoriser les « qualifiants » qui y sont associés

- *le pic du Midi d'Ossau, le gave de Pau,*
- *le lac d'Artouste et le mont d'Artouste,*
- *la ville de Saint-Jean-Pied-de-Port, La Roche-en-Ardenne*
- *la vallée à l'ouest de la ville d'Andorre-la-Vieille*
- *la ligne Bordeaux-Paris*
- *la présidence de la Communauté d'Agglomération Pau-Pyrénées*

Entité Nommée

Paik et coll. 1996 définissent 9 classes dont

Géographique : villes, ports, aéroports, îles, comtés ou départements, provinces, pays, continents, régions, fleuves, autres noms géographiques;

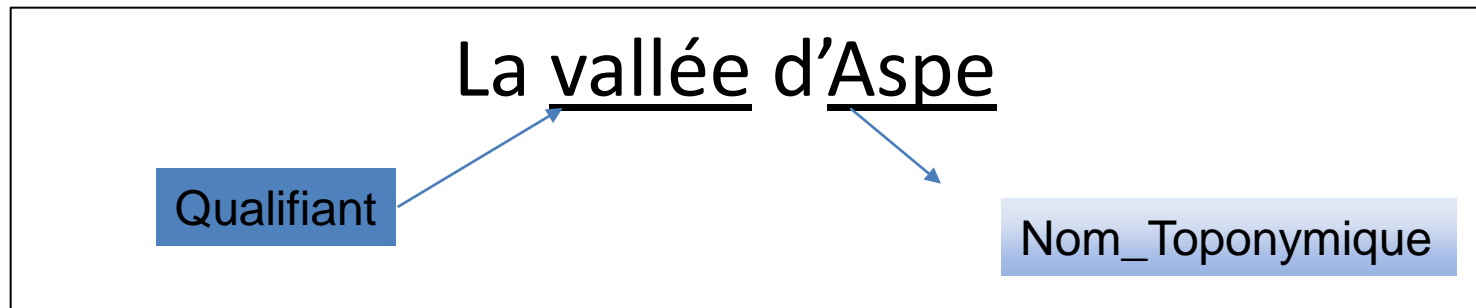
Grass 2000 définit 5 classes dont **Toponymes : noms de lieux**;

Catégorisation graphique (*Jonasson 1994*) in (*Fourour et Morin, 2003*)

- **EN pures simples** : constituées d'une seule unité lexicale commençant par une majuscule, comme *Ossau* (le plus grand nombre dans notre corpus) **(ESA)**
- **EN pures complexes** : constituées de plusieurs unités lexicales commençant par une majuscule, comme Saint-Augustin **(ESA)**
- **EN faiblement mixtes** : constituées de plusieurs mots commençant par une majuscule et contenant des mots de liaison en minuscules, comme *Bagnères de Bigorre*, *Brèche de Roland*. **(ESA)**
- **EN mixtes** : constituées de plusieurs unités lexicales dont au moins une commence par une majuscule : pas de cas recensé (*Comité international de la Croix-Rouge*)



Qualifiant et Nom Toponymique : des structures plus complexes



Nom de lieu / Toponyme

• Qualifiant peut être :

Nom commun + Adjectif

Nom commun + Adjectif + Nom commun

Nom commun = {nom commun}

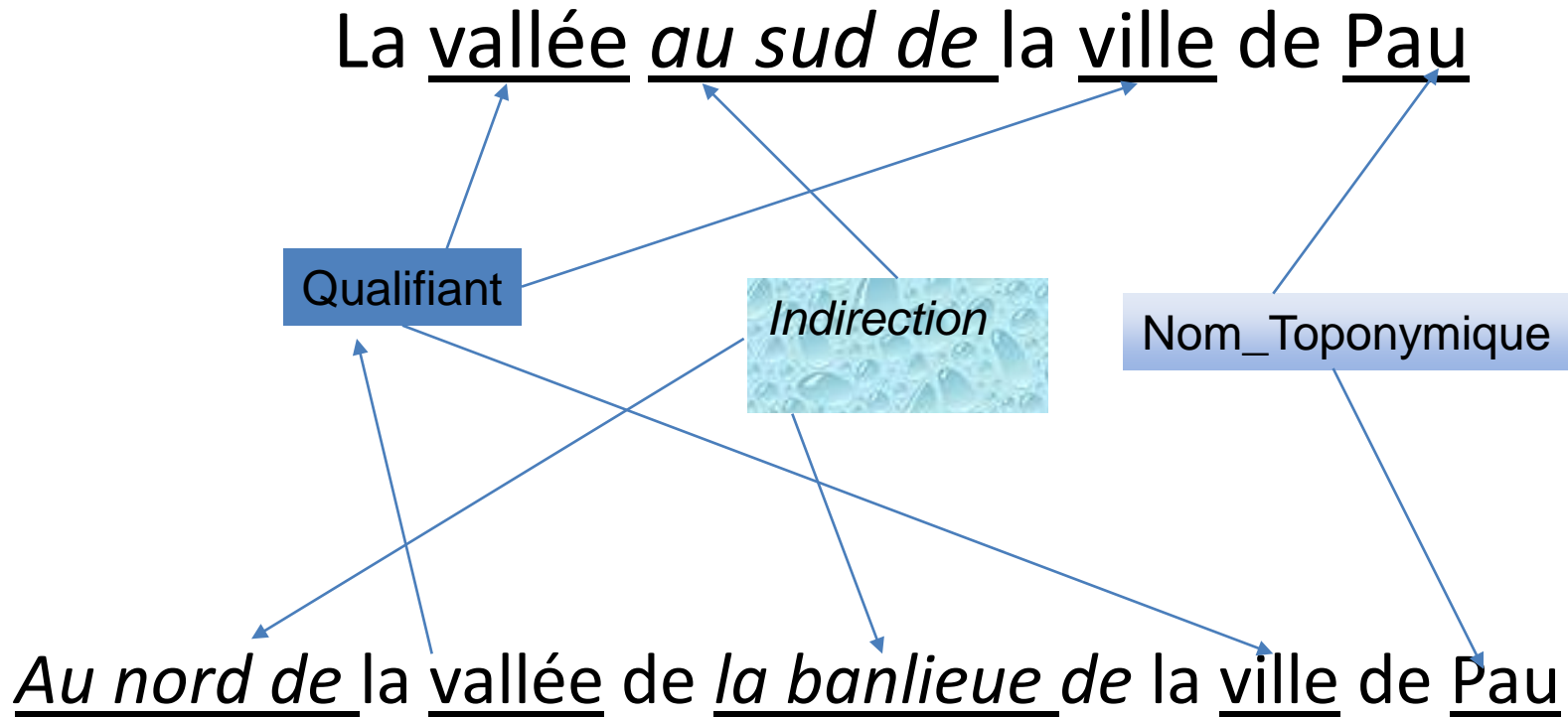
Cas particulier des *noms composés lexicaux*

À apostrophe : presqu'île, zone d'aménagement

À trait d'union : pont-tunnel, bas-fond, plate-forme insulaire, lieu-dit

Détachés (avec un espace) : hôtel de ville, aire de repos, bancs de sable, camps de plein air

Qualifiant et Nom Toponymique : la gestion des indirections



Qualifiant et Nom Toponymique : des structures plus complexes

La vallée de Mont-de-Marsan

Qualifiant

Nom_Toponymique

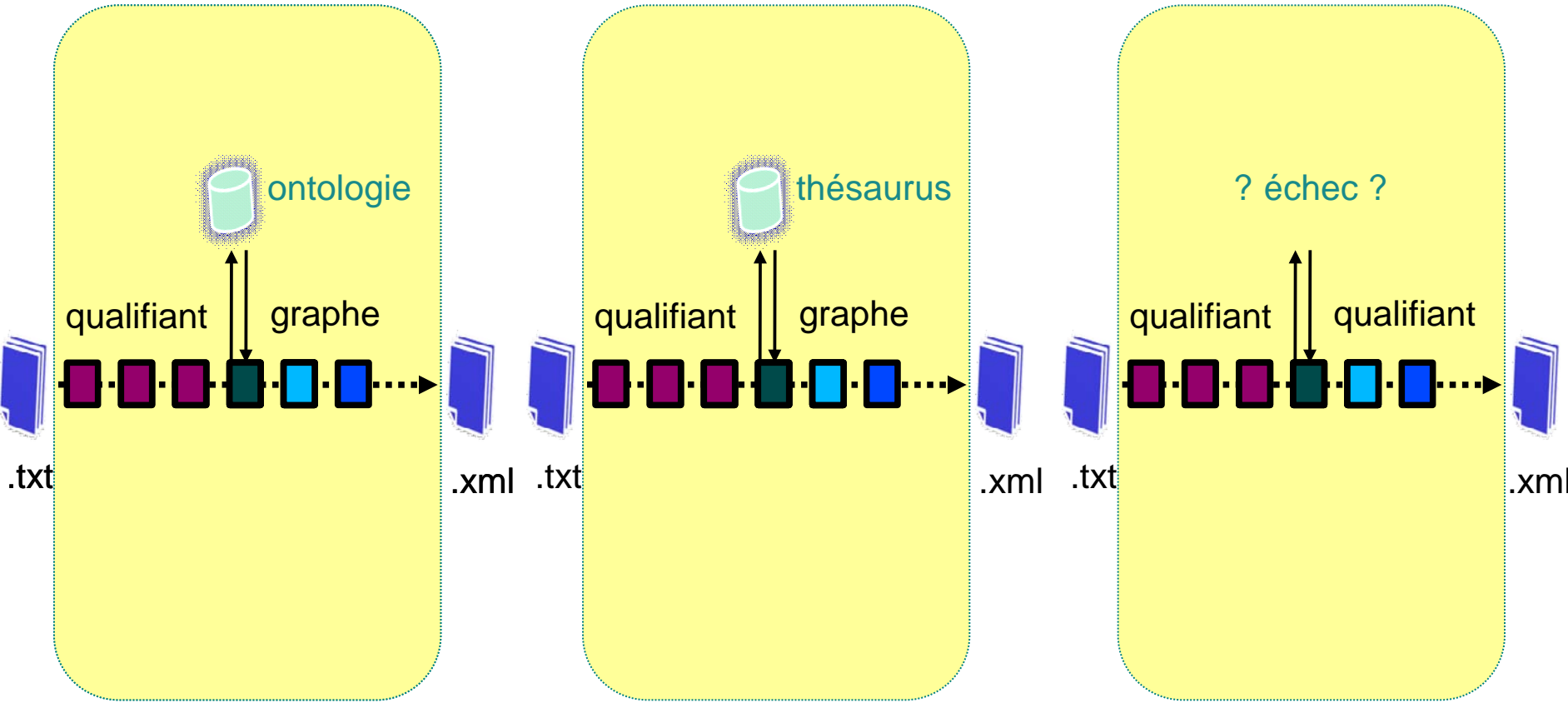


- Des Noms Toponymiques complexes :
La Roche-sur-Yon, Saint-Vincent-et-les Grenadines, l'Île Rousse,...
- Nom Toponymique :
Nom Toponymique = {Nom Toponymique , Nom Toponymique }*

Qualifiant acquérant le statut de concept



3 cas de figure



Problématique : qualifiant candidat à l'enrichissement de l'ontologie

Hypothèse : affiner la sélection des termes à analyser qui seraient candidats à l'enrichissement de l'ontologie par les patrons VP?E

« **verbe, préposition (facultative), entité spatiale** », notée sous forme d'un triplet (V,P?,E)

- Etude des relations syntagme verbal \longleftrightarrow **syntagme nominal**
[verbe de déplacement, préposition?, syntagme candidat, Entité Nommée Spatiale]
- Deux objectifs :
 - soit un terme précisant le type lors de la récupération de la géométrie du toponyme,
 - soit une proposition à l'enrichissement de l'ontologie initiale

Travaux connexes

- Patrons morphi syntaxiques (apprentissage)
- Désambiguïisation par l'emploi de ressource spécifique : Wikipédia, lexique toponymique, Geonames
- Création d'une Ressource spécifique,... (Erhman et Jacquet 2006)
- Notion de polarité aspectuelle (Boons, 1987)

- « ...Successful toponym resolution is expected to help increase precision in applications such as geographic information retrieval, topic detection and tracking and question answering....”(Leidner, 2004)

Entités
Nommées

Relations
verbales

Résolution
Géométrique

Quatre classes : personnes, lieux, organisations et expressions temporelles



Vers un typage plus fin et la désambiguïisation

Modélisation par transducteurs

Un exemple

- a) « j'ai remonté à pied la vallée d'Ossau »
- b) « j'ai marché jusqu'au pas des Echelles »
- c) « J'ai mangé mes provisions à Marsous »
- d) « J'ai marché jusqu'à la fontaine de Visos avant le déjeuner »

1. Identification des EN dans le texte et des syntagmes nominaux associés : « vallée d'Ossau », « pas des Echelles », « provisions à Marsous », « fontaine de Visos »
2. Validation des EN dans le texte : « Ossau », « Echelles » et « Marsous » sont validés mais pas Visos .
3. Identification et filtre des syntagmes nominaux liés aux EN grâce au patron (V,P ?,_,ES) : seuls a) et b) sont conservés
4. Enrichissement de l'ontologie avec les termes candidats : vallée est déjà existant dans l'ontologie initiale, pas est un terme Rameau « employé pour » (avec un sens géographique)

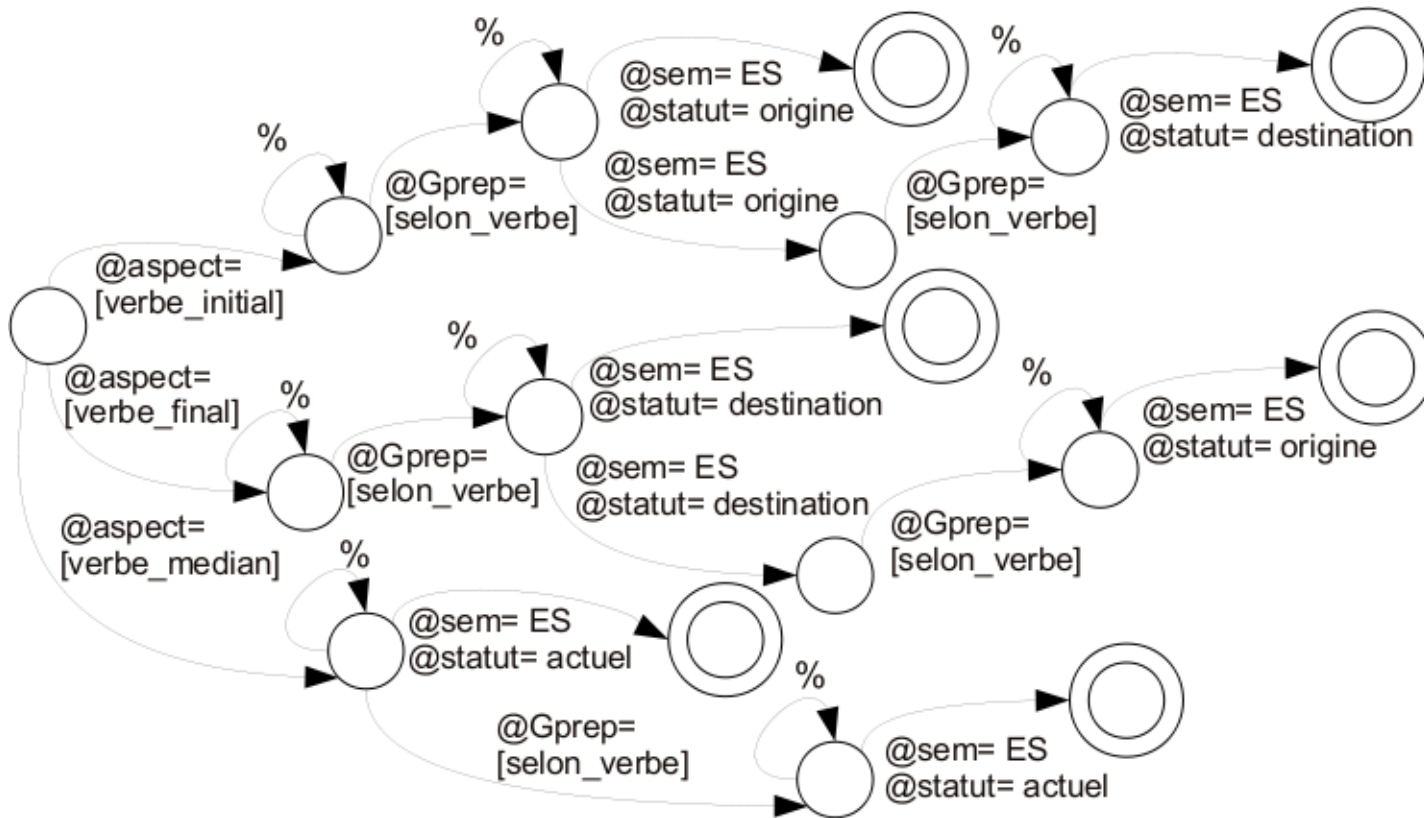
Les transducteurs sous forme de règles

3 types de polarité pour les verbes (Boons, 1987)



- « verbes initiaux » ou de polarité initiale, comme *quitter, partir, sortir, s'échapper, s'éloigner, etc.* ;
- « verbes finaux » ou de polarité finale, comme *arriver à, atteindre, entrer dans, regagner, etc.* ;
- « verbes médians » ou de polarité médiane, comme *traverser, franchir, parcourir, passer par, se déplacer dans, etc..*

Transducteur générique



Des transducteurs en DCG



Transducteurs sont sous forme de règles de grammaire dans lesquelles nous retrouvons les principaux objets du modèle : le verbe avec sa polarité, la préposition et l'ES.

Exemple d'interprétation :

« Nous arrivâmes à Crabioules »

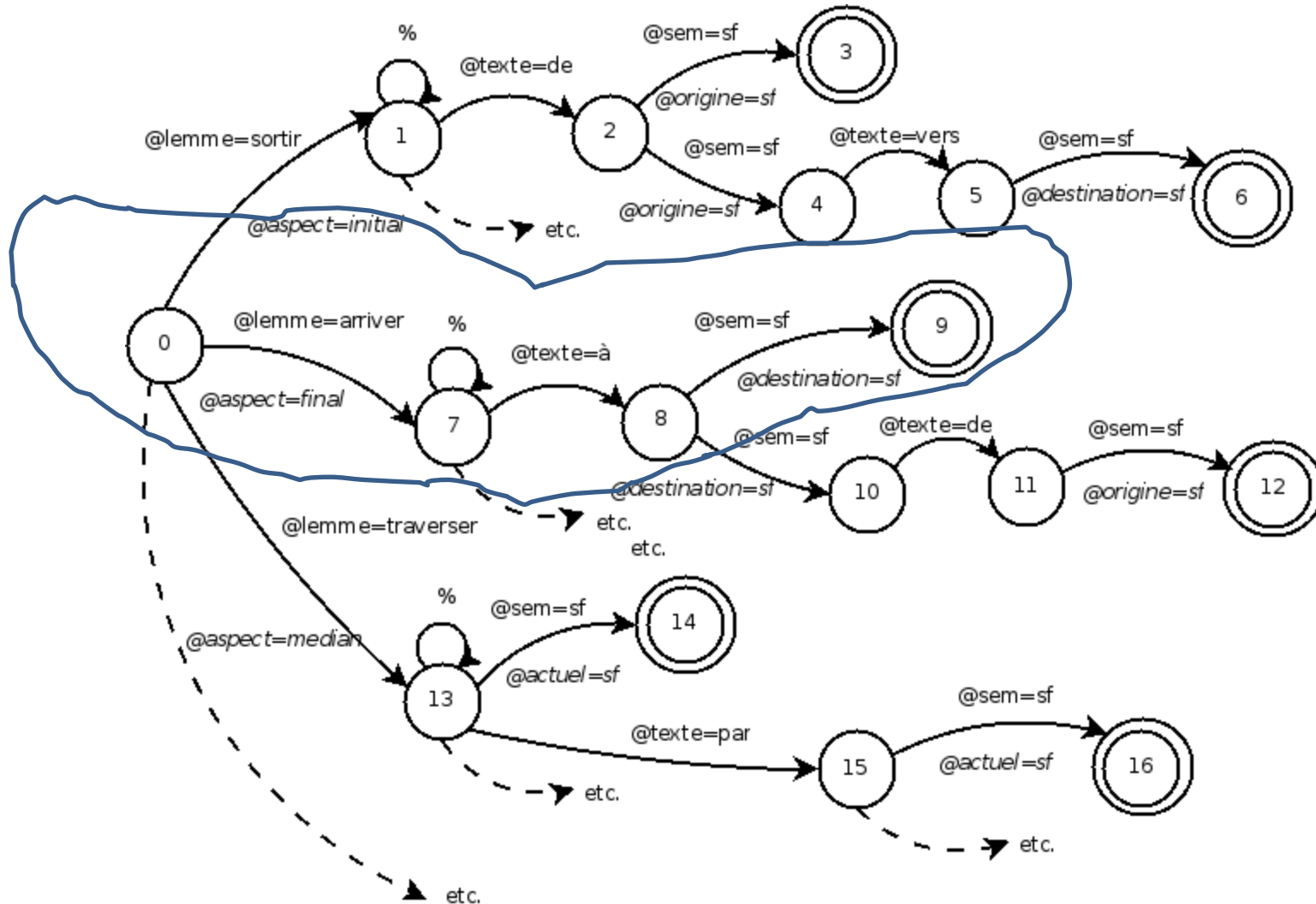
nous : @texte=nous /@lemme = nous/ @tag=pro/@stag=null/ @sem=null

arrivâmes : @texte= arrivâmes /@lemme = arriver / @tag=ver/@stag=ppa/
@sem=null

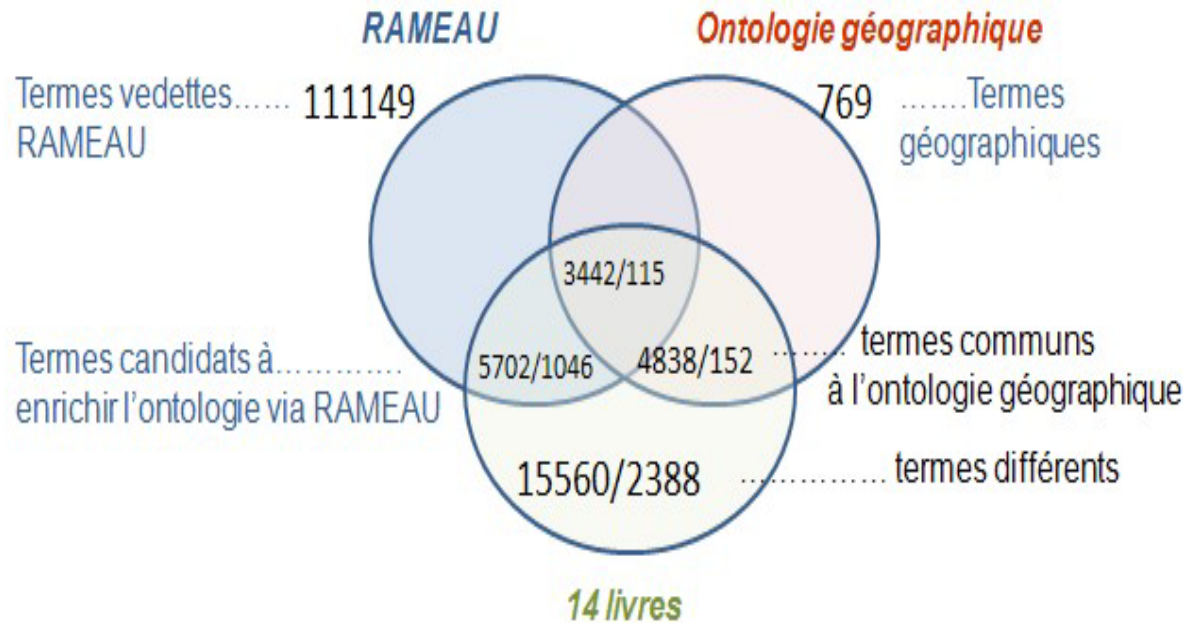
à : @texte= à /@lemme = à/ @tag=pre/@stag=null/ @sem=null

Crabioules : @texte= Crabioules /@lemme = Crabioules /
@tag=nom/@stag=prp/ @sem=es

Grâce à ce transducteur...

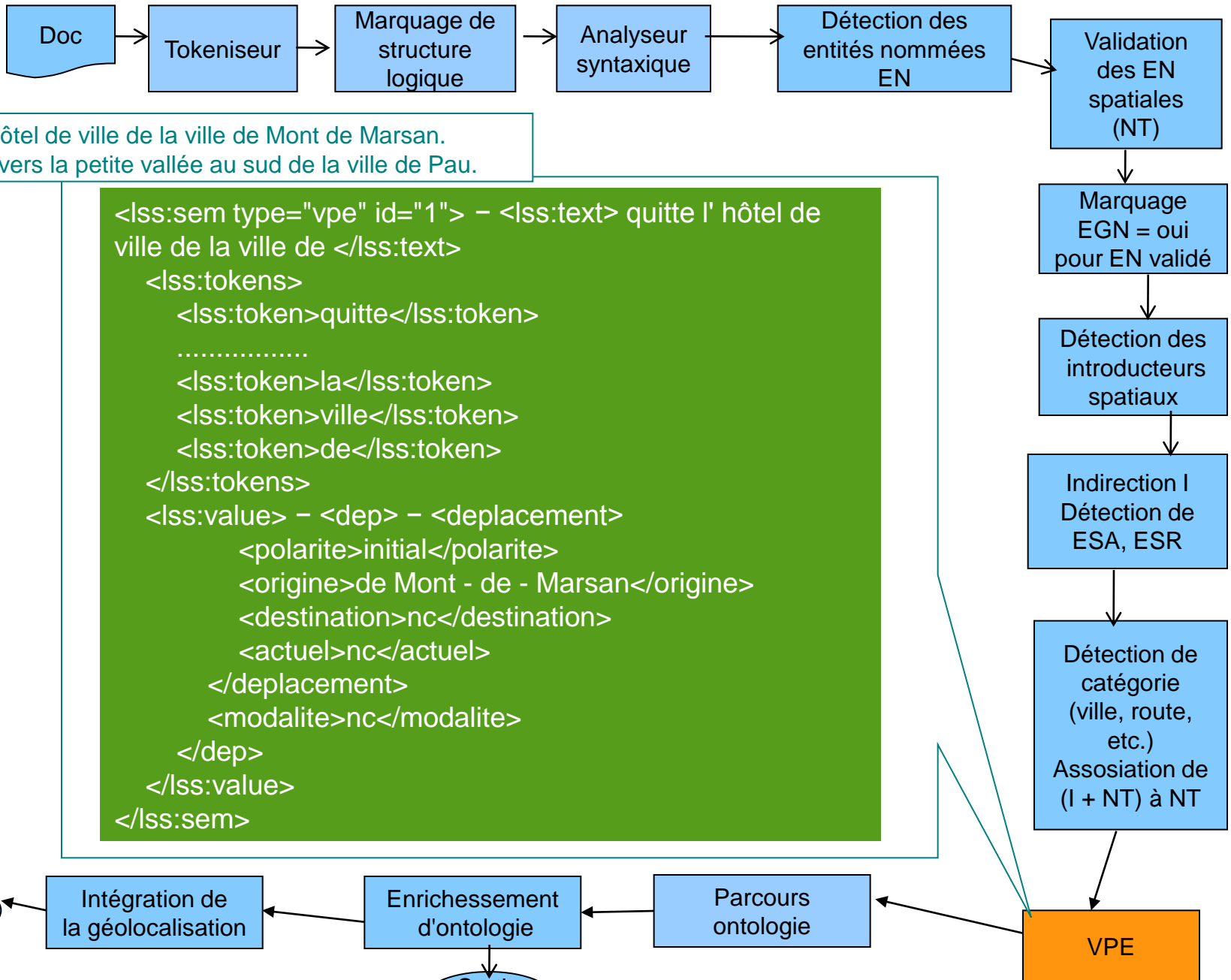


Estimations



Toponymes candidats & termes associés.

- 1046 termes RAMEAU sont candidats à l'enrichissement de notre ontologie.
- Grâce au filtre (V,P ?,_,ES), on réduit ce lexique à 153 termes Rameau et 140 termes de source inconnue.



Extrait fichier SKOS produit



```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://geonto.lri.fr/skos/ontologyEnrichment.rdf#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:owl2="http://www.w3.org/2006/12/owl2#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xml:base="http://geonto.lri.fr/skos/ontologyEnrichment.rdf">
  <owl:Ontology rdf:about=""/>
  .....
  // Individuals
  //
  -->
  <!-- http://geonto.lri.fr/skos/ontologyEnrichment.rdf#charmante_vall&#233;e -->
  <skos:Concept rdf:about="#charmante_vallée">
    <skos:prefLabel xml:lang="fr">charmante vallée</skos:prefLabel>
  </skos:Concept>
  <!-- http://geonto.lri.fr/skos/ontologyEnrichment.rdf#coueyla -->
  <skos:Concept rdf:about="#coueyla">
    <skos:prefLabel xml:lang="fr">coueyla</skos:prefLabel>
  </skos:Concept>
  <!-- http://geonto.lri.fr/skos/ontologyEnrichment.rdf#poste -->
  <skos:Concept rdf:about="#poste">
    <skos:prefLabel xml:lang="fr">poste</skos:prefLabel>
  </skos:Concept>
</rdf:RDF>

```

Limites actuelles



- Des Constructions [*verbe de déplacement, préposition (facultative), syntagme nominal, entité nommée spatiale*] qui posent problème
 - (a) « j'ai marché jusqu'au maire d'Accous »,
 - (b) « Je pus pénétrer dans les solitudes des Landes » et
 - (c) « Je suis parti à cheval pour Gavarnie »

(a) maire ? (source : Rameau)

(b) solitudes? (source : inconnue)

Sens géographique ?

- (c) « à cheval » : nécessité de mettre en place un patron lexico-syntaxique identifiant ces syntagmes comme une **modalité de déplacement**

Objectif relation VPE

Au final nous assignons un double objectif à la relation VPE pour les qualifiants.

Soit Q un qualifiant

1. Cas démontré et mis en place dans notre chaîne de traitement

SI Q est associé à un Nom Propre validé comme toponyme

SI Q ontologie de référence ALORS

Q marqué comme étant probablement un concept géographique et donc un "bon" candidat à l'enrichissement;

Q ou thésaurus RAMEAU

Objectif relation VPE

Au final nous assignons un double objectif à la relation VPE pour les qualifiants.

Soit Q un qualifiant

1. Cas à mettre en place dans notre chaîne de traitement

SI Q est associé à un Nom Propre non validé comme toponyme

SI Q ∈ ontologie de référence ALORS

Nom Propre marqué comme étant probablement un candidat à devenir
Toponyme

Exemple (vu en T10):

d) « J'ai marché jusqu'à la fontaine de Visos avant le déjeuner »

Des questions?



pour votre attention

