EGC 2026 - Anglet

29 janvier 2026

Construction automatique d’un graphe de connaissances géo-historiques à partir de textes encyclopédiques anciens

Bin Yang, Ludovic Moncla, Fabien Duchateau, Frédérique Laforest

Les articles de lieux dans l’Encyclopédie

Jean-pied-de-Port, S. (Géog.) ville de France en Gascogne, à une lieue des frontieres d'Espagne, autrefois capitale de la basse Navarre, [...] elle est sur la Nive, à l'entrée d'un des passages des Pyrénées, à 8 lieues S. E. de Bayonne, 12 N. E. de Pampelune, 176 S. O. de Paris. Long. 16. 22. lat. 43. 8. (D. J.)

Jean de Luz, S. (Géog.) Lucius Vicus ; le nom basque est Loitzun, petite ville de France en Gascogne, la deuxieme du pays de Labour, & la derniere du côté de l'Espagne, avec un port. Elle est sur une petite riviere, que Piganiol de la Force nomme la Ninette, & M. de Lisle le Nivelet, à 4 lieues N. E. de Fontarabie, 4 S. O. de Bayonne, 174 S. O. de Paris. Long. 15. 59. 28. lat. 43. 23. 15. (D. J.)

PYRÉNÉES, les (Géogr. anc.) Pyrenaei montes ; montagnes d'Europe aux frontieres de la France & de l'Espagne [...]
Elles commencent au port de Vendres dans le Roussillon, sur la Méditerranée, & à Saint-Jean-de-Luz dans la Biscaye françoise, sur l'Océan, d'où elles s'étendent jusqu'à Saint-Sébastien, port de mer dans la Biscaye espagnole, à Pampelune dans la Navarre, à Venasca dans l'Arragon, à Lérida & à Tortose, dans la Catalogne. [...]

GASCOGNE, (la) Vasconia, (Géog.) grande province de France qui fait une partie du gouvernement général de la Guienne ; elle est entre la Garonne, l'Océan, & les Pyrénées [...]

Environ 15 500 articles de Géographie (parmi 75 000), dans lesquelles environ 15 000 décrivent des lieux

Pas d’index de lieux, les coordonnées sont exprimées dans le texte (pour 4 798 articles).

Les articles de lieux dans l'Encyclopédie

Jean-pied-de-Port, S. (Géog.) ville de France en Gascogne, à une lieue des frontieres d'Espagne, autrefois capitale de la basse Navarre, [...] elle est sur la Nive, à l'entrée d'un des passages des Pyrénées, à 8 lieues S. E. de Bayonne, 12 N. E. de Pampelune, 176 S. O. de Paris. Long. 16. 22. lat. 43. 8. (D. J.)

Jean de Luz, S. (Géog.) Lucius Vicus ; le nom basque est Loitzun, petite ville de France en Gascogne, la deuxieme du pays de Labour, & la derniere du côté de l'Espagne, avec un port. Elle est sur une petite riviere, que Piganiol de la Force nomme la Ninette, & M. de Lisle le Nivelet, à 4 lieues N. E. de Fontarabie, 4 S. O. de Bayonne, 174 S. O. de Paris. Long. 15. 59. 28. lat. 43. 23. 15. (D. J.)

PYRÉNÉES, les (Géogr. anc.) Pyrenaei montes ; montagnes d'Europe aux frontieres de la France & de l'Espagne [...]
Elles commencent au port de Vendres dans le Roussillon, sur la Méditerranée, & à Saint-Jean-de-Luz dans la Biscaye françoise, sur l'Océan, d'où elles s'étendent jusqu'à Saint-Sébastien, port de mer dans la Biscaye espagnole, à Pampelune dans la Navarre, à Venasca dans l'Arragon, à Lérida & à Tortose, dans la Catalogne. [...]

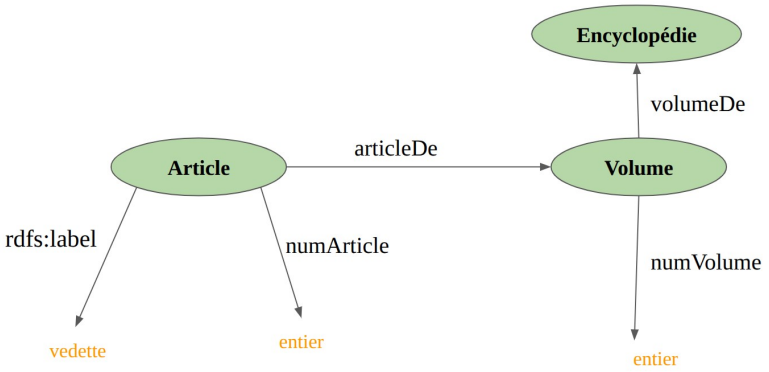
GASCOGNE, (la) Vasconia, (Géog.) grande province de France qui fait une partie du gouvernement général de la Guienne ; elle est entre la Garonne, l'Océan, & les Pyrénées [...]

Environ 15 500 articles de Géographie (parmi 75 000), dans lesquelles environ 15 000 décrivent des lieux

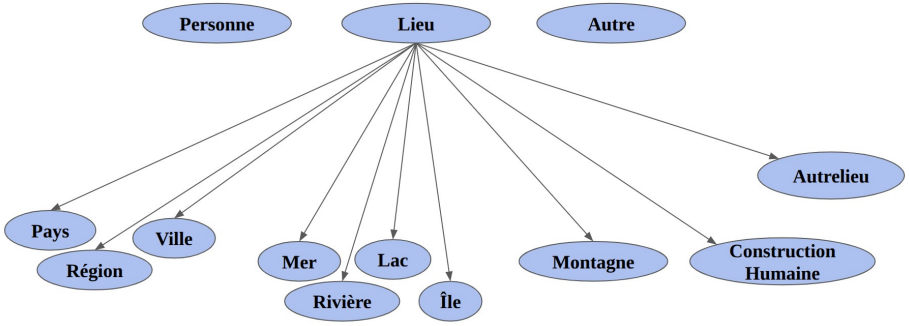
Pas d'index de lieux, les coordonnées sont exprimées dans le texte (pour 4 798 articles).

Modélisation

Ontologie Provenance : structure du corpus



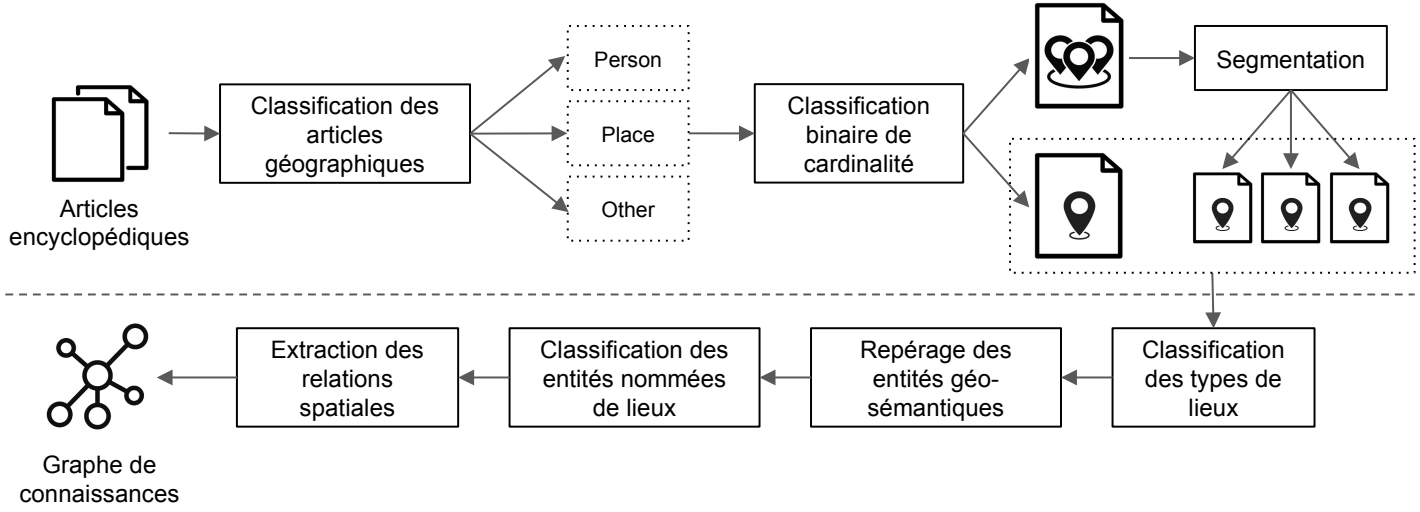
Ontologie Spatiale : concepts géographiques & relations et propriétés spatiales



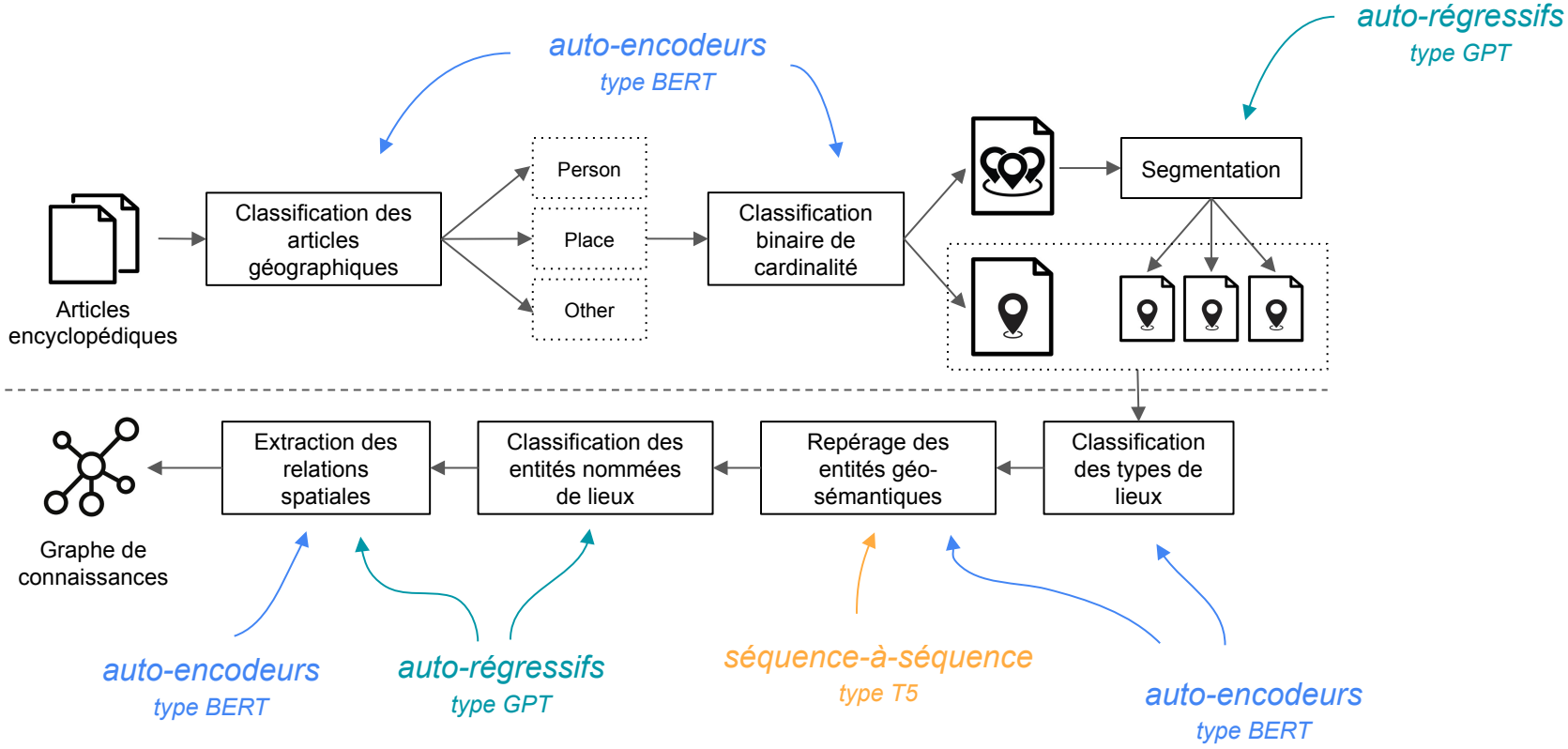
Inclusion: “dans”, “ville de”…
Adjacence: “à côté de”, “est adjacent à”…
Orientation: “au sud de”…
Distance: “à 20 lieues de”…
Mouvement: “se jette dans”…
Crosses: “sur la rivière”, “traverse la ville”…
Autre_relation: “par”, “au dessous de”…

a_longueur: “2km”…
a_surface: “25 kilo mètres carés”…
a_longitude: “Long. 23.31.”…
a_latitude: “Lat. 45.11.”…

Chaîne de traitements pour la construction automatique d'un graphe de connaissances



Chaîne de traitements pour la construction automatique d'un graphe de connaissances



De l'IA et des *Transformers* à toutes les étapes !

Illustration de la chaîne de traitement

PARIS, *ville capitale du royaume de France, sur la Seine, 90 lieues sud-est de Londres…*

FRANCE, *grand pays en Europe*

LYONNAIS, *habitants de lyon,...*

Latitude, *terme géographique…*

VERA, *(Géog. anc.),*
1° nom d'une ville de Médie, selon [...]
2° nom d'un fleuve de la Gaule…

Illustration de la chaîne de traitement

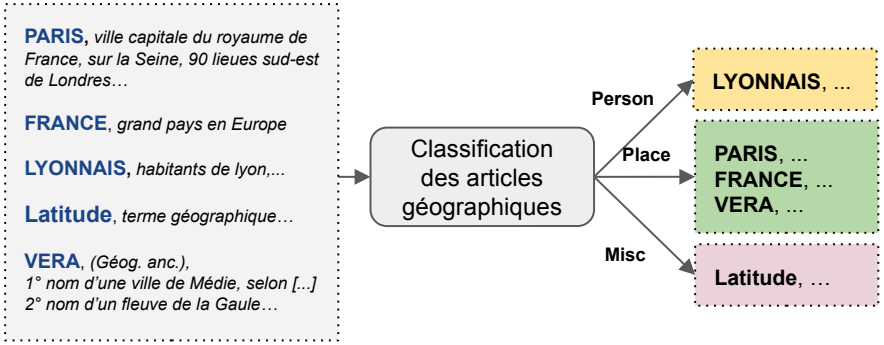


Illustration de la chaîne de traitement

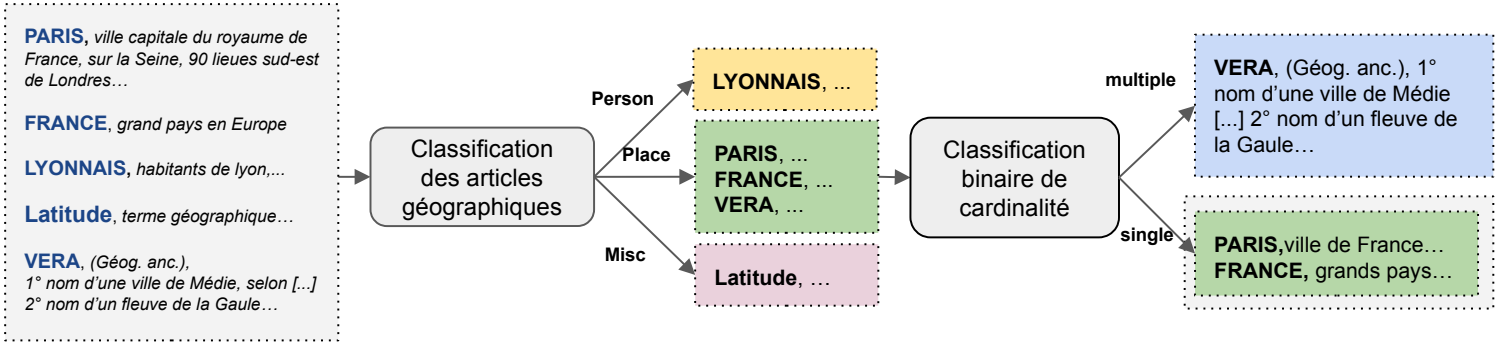


Illustration de la chaîne de traitement

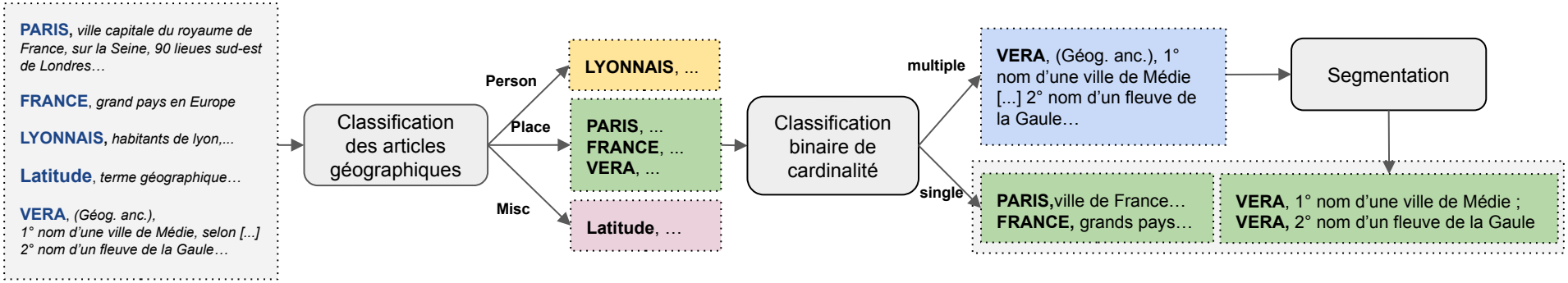


Illustration de la chaîne de traitement

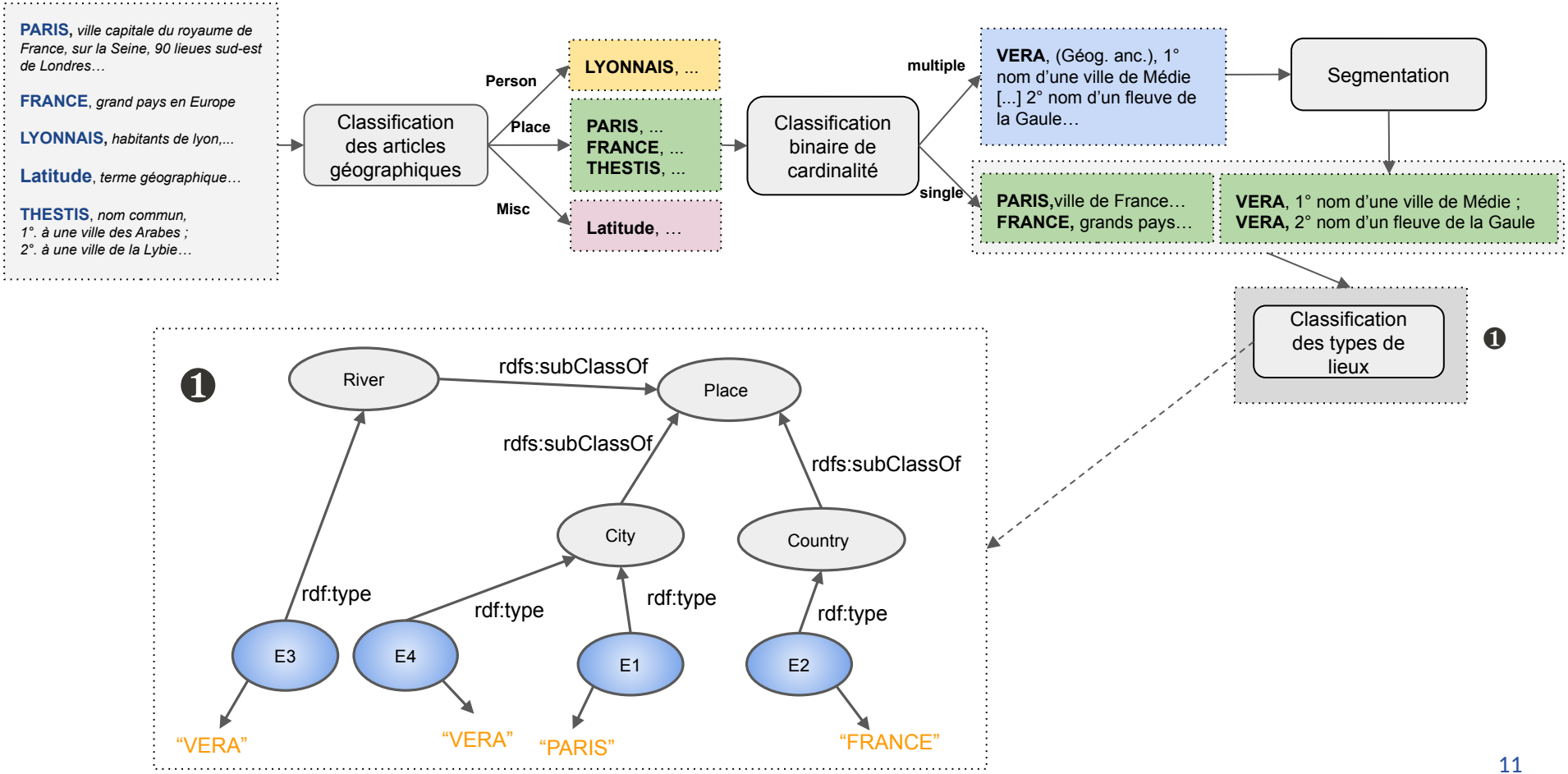


Illustration de la chaîne de traitement

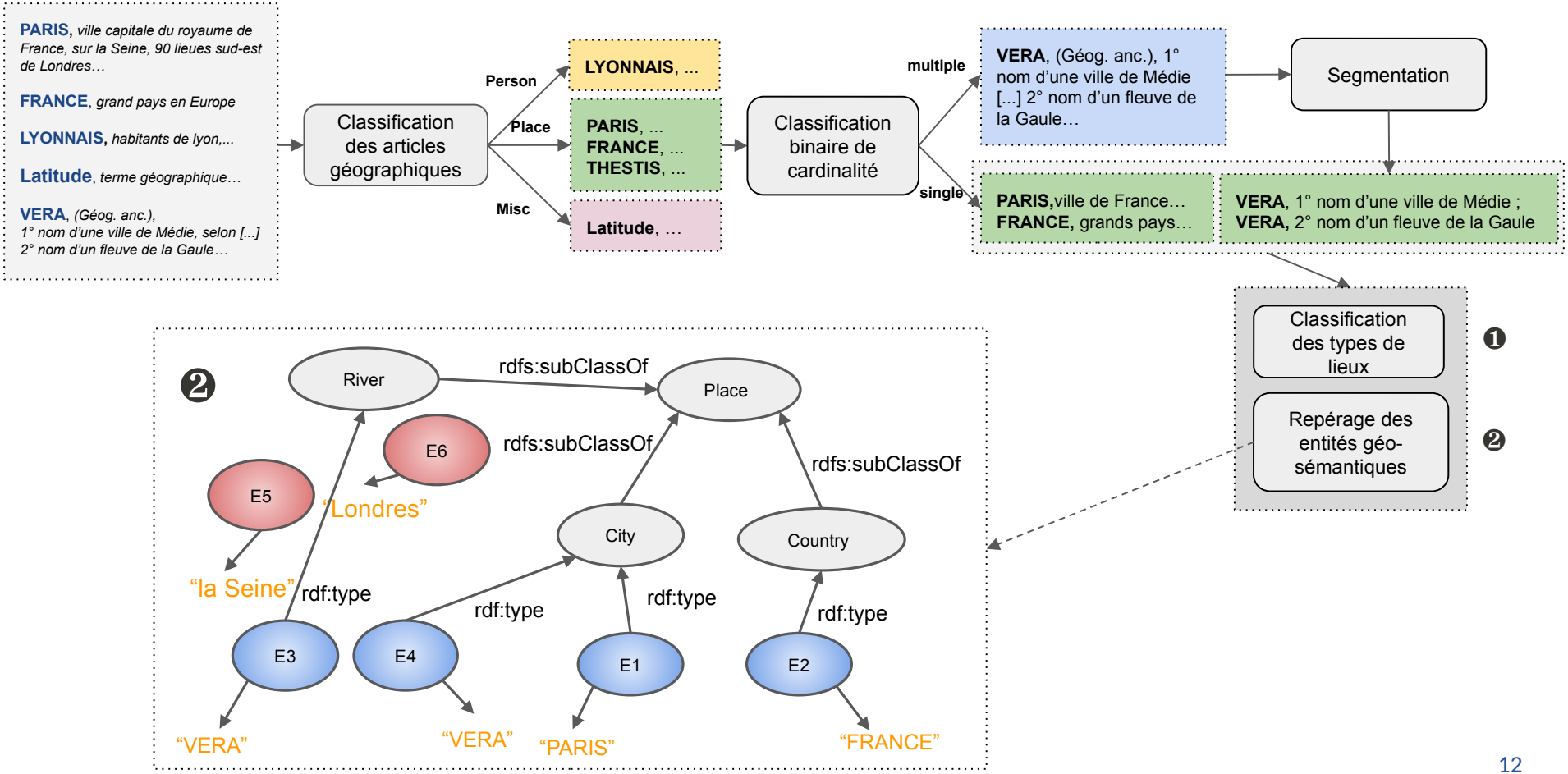


Illustration de la chaîne de traitement

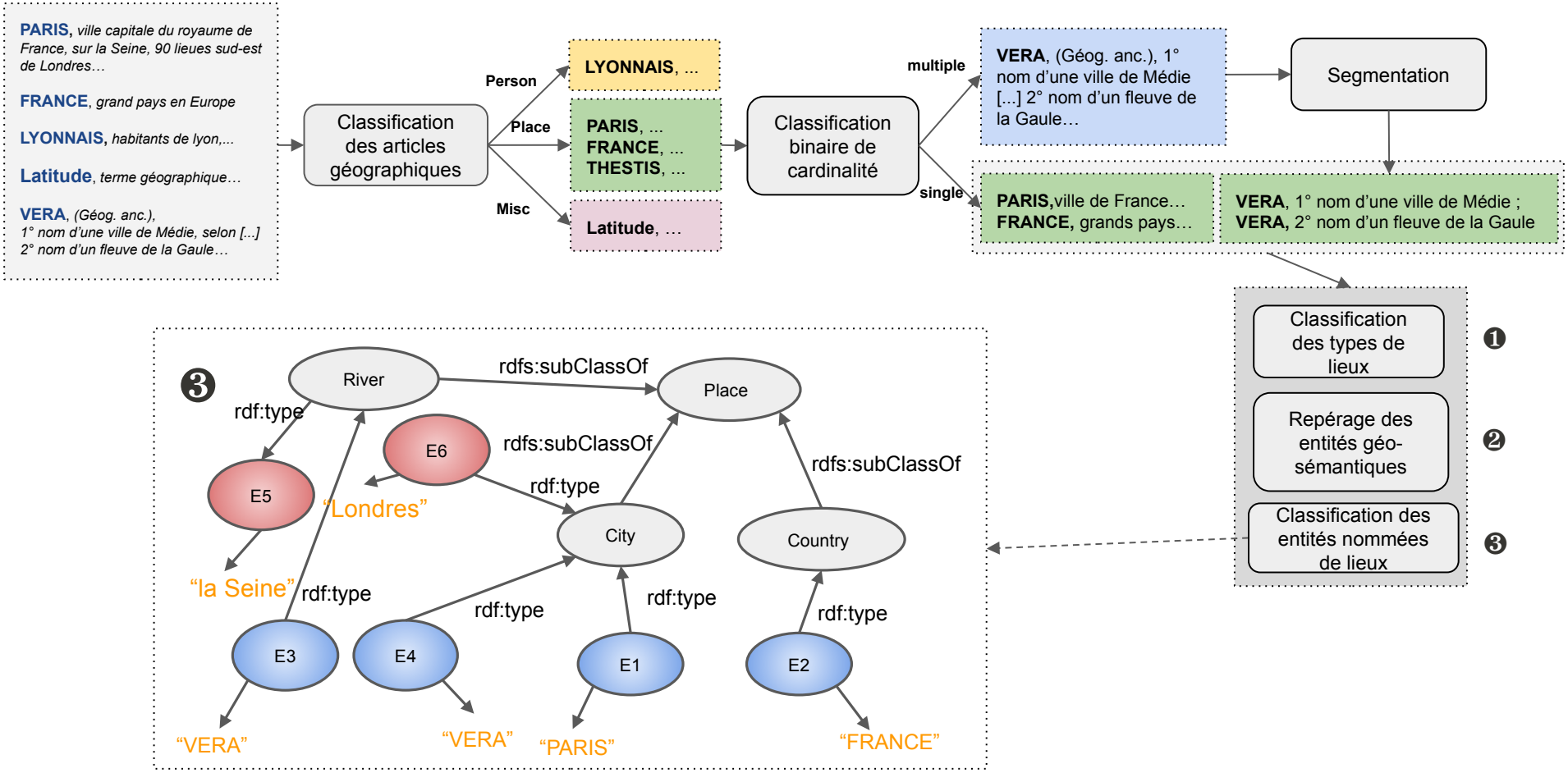
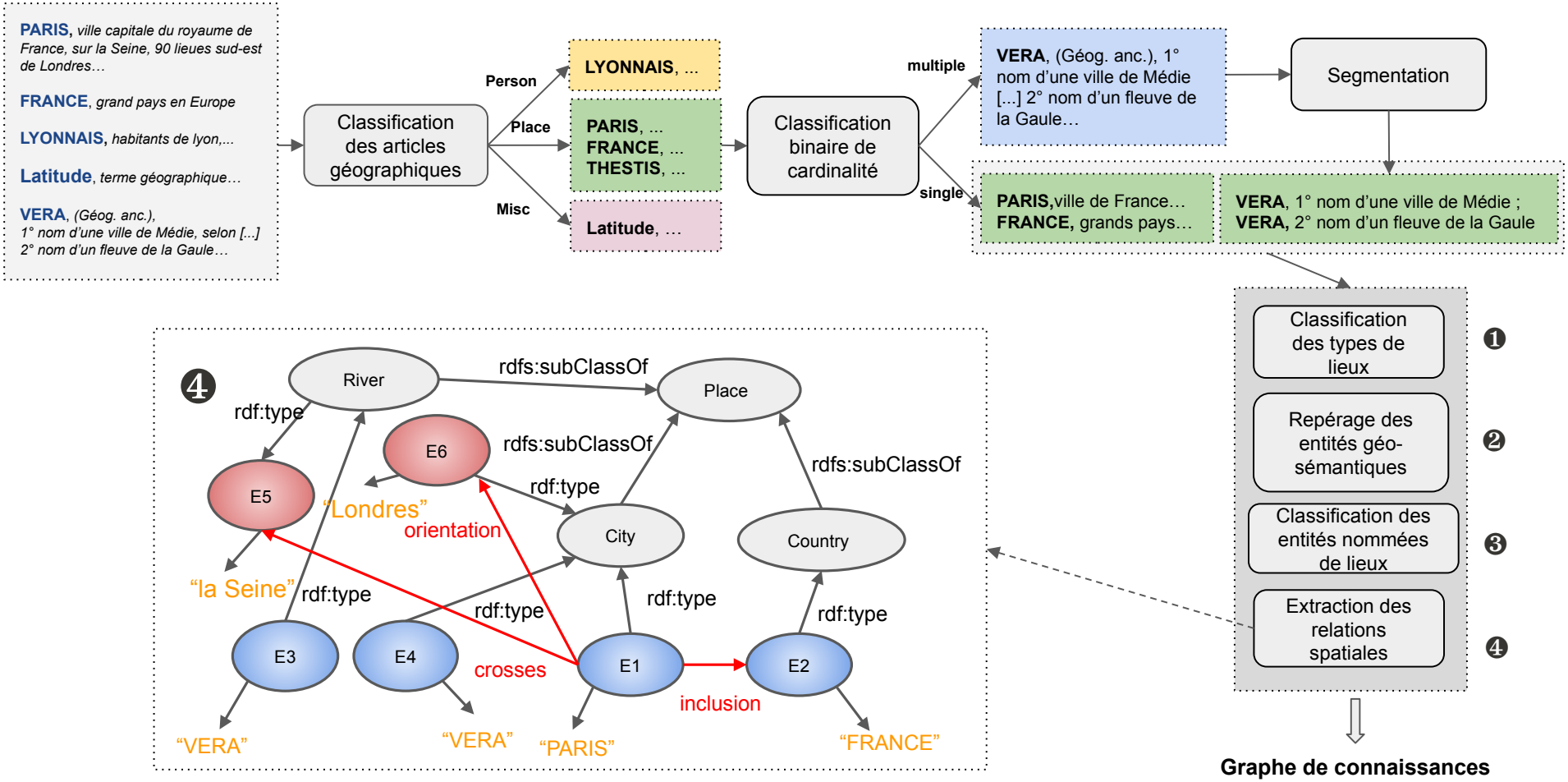
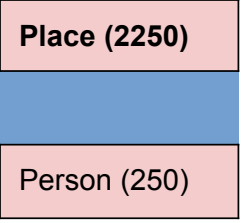


Illustration de la chaîne de traitement

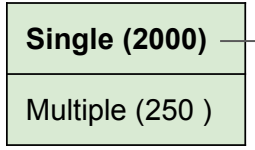


données pour l’entraînement des modèles



Type d’article de Géographie

Cardinalité



Type de lieu

1 012 City	**59** HumanMade
263 Island	**51** Sea
188 Region	**51** Other
181 River	**43** Country
114 Mountain	**40** Lake

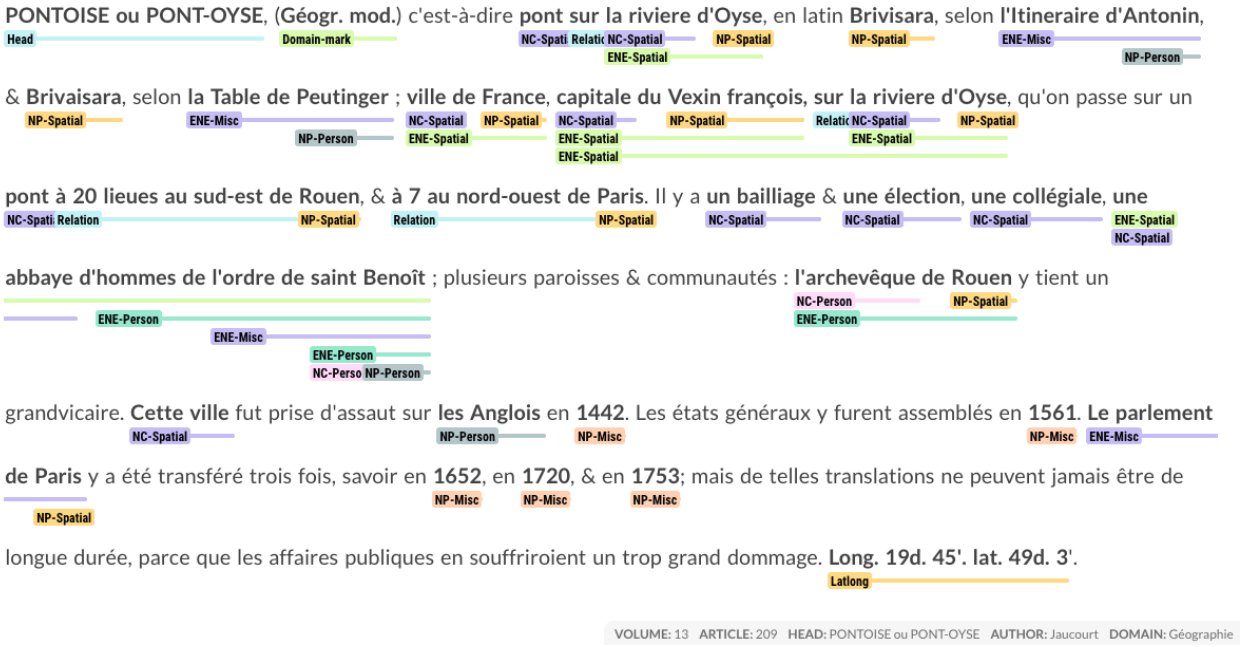
Place	Single	City	TUTUCURIN, TUTOCORIN, TUTUCORY, (Géog. mod.) ville de la presqu'isle occidentale de…
Place	Single	City	KARAHISAR, (Géog.) ville détruite de la Natolie, qui est, selon Paul Lucas, dans so…
Place	Single	Region	PHTHIOTIDE, (Géog. anc.) Phthiotis ; province de la Thessalie. Ptolomée y place…
Place	Single	City	ORFA, (Géogr.) M. de Lisle dit Ourfa, ville d'Asie à l'Orient de l'Euphrate dans le…
Person	null	null	* Ariens, s. m. pl. (hist. & Géog.) peuples d'Allemagne, dont Tacite fait mention, & qu…
Place	Single	Island	MALTHE, (Géog.) en grec Μελίτη, en latin Melita, île de la mer Méditerranée, entre…
Place	Multiple	null	JAVA, (l'Isle de) Géog. nom de deux îles de la mer des Indes, dont l'une est appellée l…
Person	null	null	LEVI, ou LEVÉ, (Géog. anc.) & par Polybe, l. II. c. xvij. Λάοι, Laoi, ancien peuple…
Other	null	null	Emprunt, terme de Riviere, se dit d'un passage qui mene à la travure d'un bateau…
Place	Single	River	WENSBEEK, le, (Géog. mod.) en latin Venta, petite riviere d'Angleterre. Elle prend sa…

GeoEDdA-TopoRel : https://huggingface.co/datasets/GEODE/GeoEDdA-TopoRel

Jeu de données annoté en entités sémantiques

GeoEDdA-NER dataset

- 25 087 spans annotés
- annotation semi-manuelle



Moncla, L., Vigier, D., & McDonough, K. (2024). *GeoEDdA: A Gold Standard Dataset for Named Entity Recognition and Span Categorization Annotations of Diderot & d'Alembert's Encyclopédie*. Proceedings of the 2nd International Workshop on Geographic Information Extraction from Texts (GeoExT). Glasgow, Scotland.
https://huggingface.co/datasets/GEODE/GeoEDdA-NER

Classification des types de lieux (articles de Géographie)

- BERT-based fine-tuning (text classification)

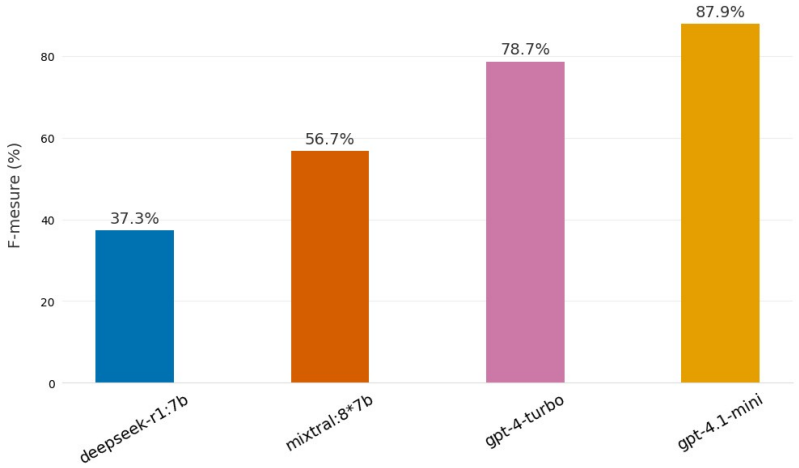
ZNAIM ou ZNOYM, (Géog. mod.) ville de Boheme, en Moravie, sur la Teya, vers les…	City
SABRATA, (Géogr. anc.) Sabrata colonia, ville maritime & colonie romaine en Afrique…	City
Cuba, (Géog. mod.) grande île de l'Amérique septentrionale, à l'entrée du golfe du…	Island
DIVANDUROU, (Géog. mod.) nom de cinq îles d'Asie, voisines des Maldives.	Island
VIVERO, ou BIVERO, (Géog. mod.) petite ville d'Espagne, dans la Galice, lur une montagne…	City
SENSET, le, ou la SANSSE, (Géog. mod.) petite riviere des Pays-bas ; elle prend sa…	River
MENOVIA, (Géog. anc.) ancienne ville d'Angleterre avec évêché suffragant de…	City
CARIBANE, (Géog.) province maritime de l'Amérique méridionale, qui s'étend depuis…	Region

	Précision	Rappel	F-mesure
City	90.91%	100%	95.24%
Island	96.30%	96.30%	96.30%
Region	89.66%	92.86%	91.23%
River	96.56%	100%	98.25%
Montagne	100%	95.45%	97.67%
HumanMade	90.00%	100%	94.74%
Other	88.89%	66.67%	76.19%
Sea	100%	91.67%	94.74%
Lake	100%	88.89%	94.12%
Country	100%	92.31%	96.00%
average	**94.63%**	**94.50%**	**94.37%**

Précision, Rappel et F-mesure de la classification de type des vedettes

Classification des entités nommées de lieux

- Few-shot prompting



F-mesure de la classification des entités nommées de lieux par LLMs

- BERT-based fine-tuning

	Précision	Rappel	F-mesure	accuracy average
4-grammes précédents	82.53%	82.10%	82.03%	82.10%
4-grammes précédents et suivants	83.99%	83.32%	83.28%	83.32%
5-grammes précédents et suivants	85.12%	84.85%	83.80%	84.73%
8-grammes précédents et suivants	85.16%	84.54%	84.49%	84.54%

Précision, Rappel et F-mesure de la classification des entités nommées de lieux selon la taille de contexte

Extraction et normalisation des coordonnées

- Normalisation format DMS
- Modèle mT5 (séquence à séquence)

		EM	CER	Support
	mt5-small	0.86	0.07	959
	gpt5-mini	0.86	**0.03**	
points only	mt5-small	**0.92**	0.01	903
	gpt5-mini	0.90	0.01	
others only	mt5-small	0.00	0.57	56
	gpt5-mini	**0.29**	**0.22**	

* AIX, (Géog.) ville de France en Provence, dont elle est la capitale, près de la petite riviere d'Arc. Long.…	[["43 31' 35\" N 23 6' 34\" E"]]
* Aix, (Géog.) ville de Savoye sur le lac de Bourget. Long. 23. 34. lat. 45. 40.	[["45 40' N 23 34' E"]]
* Aix-la-Chapelle, (Géog.) ville d'Allemagne dans le cercle de Westphalie au Duché de Juliers. Long. 23. 55. lat. 51.…	[["51 55' N 23 55' E"]]
* AKISSAR ou AK-HISSAR, (Géog.) ville d'Asie dans la Natolie, sur la riviere Hermus. Long. 46. lat. 38. 50.	[["38 50' N 46 E"]]
* Alais, (Géog.) ville de France dans le bas Languedoc sur la riviere de Gardon. Long. 21. 32. lat. 44. 8.	[["44 8' N 21 32' E"]]
* ALATRI, (Géog.) ancienne ville d'Italie, dans la campagne de Rome. Long. 30. 58 lat. 41. 44.	[["41 44' N 30 58' E"]]
* ALBANIE, (Geog.) province de la Turquie Européenne sur le golphe de Venise. Long. 36. 18-39. 40. lat. 39-43. 30.	[["39 N 36 18' E", "43 30' N 39 40' E"]]
* ALBANO, (Géog.) ville d'Italie sur un lac de même nom, dans la campagne de Rome. Long. 30. 15. lat. 41. 43.	[["41 43' N 30 15' E"]]
* ALBANOPOLI, (Géog.) ville de la Turquie Européenne dans l'Albanie. Long. 38. 4. lat. 51. 48.	[["51 48' N 38 4' E"]]

https://huggingface.co/GEODE/mt5-small-coords-norm

Classification des relations et *relation-linking*

- Classification des relations spatiales
  - BERT-based fine-tuning

	Précision	Rappel	F-mesure
Inclusion	98.06%	91.18%	97.74%
Distance-Orientation	93.44%	99.13%	96.20%
Adjacence	95.71%	85.90%	90.54%
Crosses	76.45%	100%	86.67%
Mouvement	96.88%	91.18%	93.94%
Autre-relation	97.43%	88.37%	92.68%
	95.09%	**94.69%**	**94.69%**

distance: *"90 lieues de"*

orientation: *"à l'est de"*

distance-orientation: *"trois lieues au sud de"*

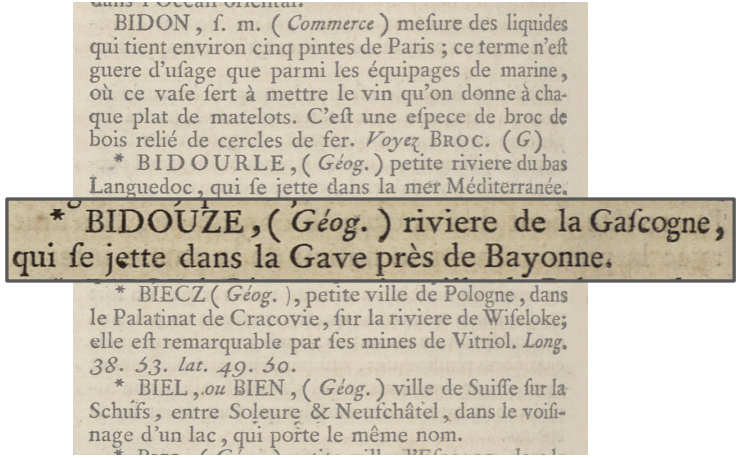
Regex

distance ou/et orientation

- Relation-linking
  - Few-shot prompting
  - Approche probabiliste

	nb_correct	nb_total	précision
llama3 :70b	106	124	85.48%
gpt-4-turbo	113	124	91.12%
gpt-4.1-mini	113	124	91.12%
approche probabiliste	115	124	92.74%

Illustration du workflow d’extraction d’information



BIDON, ſ. m. (*Commerce*) meſure des liquides qui tient environ cinq pintes de Paris ; ce terme n’eſt guere d’uſage que parmi les équipages de marine, où ce vaſe ſert à mettre le vin qu’on donne à chaque plat de matelots. C’eſt une eſpece de broc de bois relié de cercles de fer. *Voyez* BROC. (*G*)

* BIDOURLE, (*Géog.*) petite riviere du bas Languedoc, qui ſe jette dans la mer Méditerranée.

* BIDOUZE, (*Géog.*) riviere de la Gaſcogne, qui ſe jette dans la Gave près de Bayonne.

* BIECZ (*Géog.*), petite ville de Pologne, dans le Palatinat de Cracovie, ſur la riviere de Wiſeloke; elle eſt remarquable par ſes mines de Vitriol. *Long.* 38. 53. *lat.* 49. 50.

* BIEL, *ou* BIEN, (*Géog.*) ville de Suiſſe ſur la Schuſs, entre Soleure & Neufchâtel, dans le voiſinage d’un lac, qui porte le même nom.

Vol. 2, page 242

Illustration du workflow d’extraction d’information

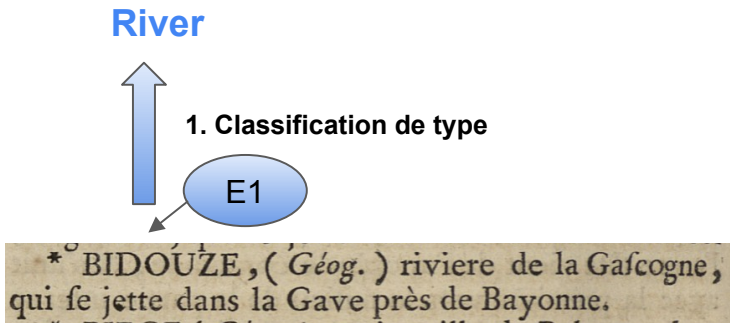


Illustration du workflow d’extraction d’information

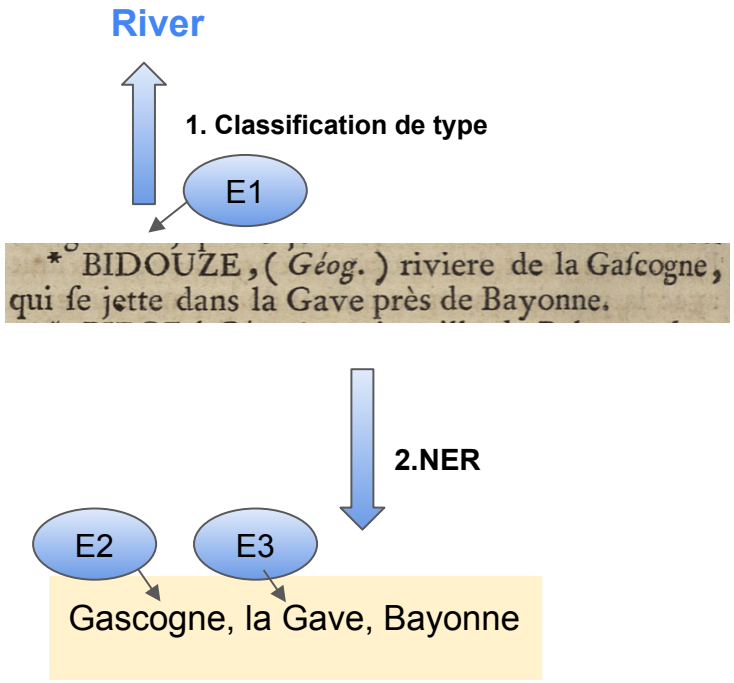


Illustration du workflow d'extraction d'information

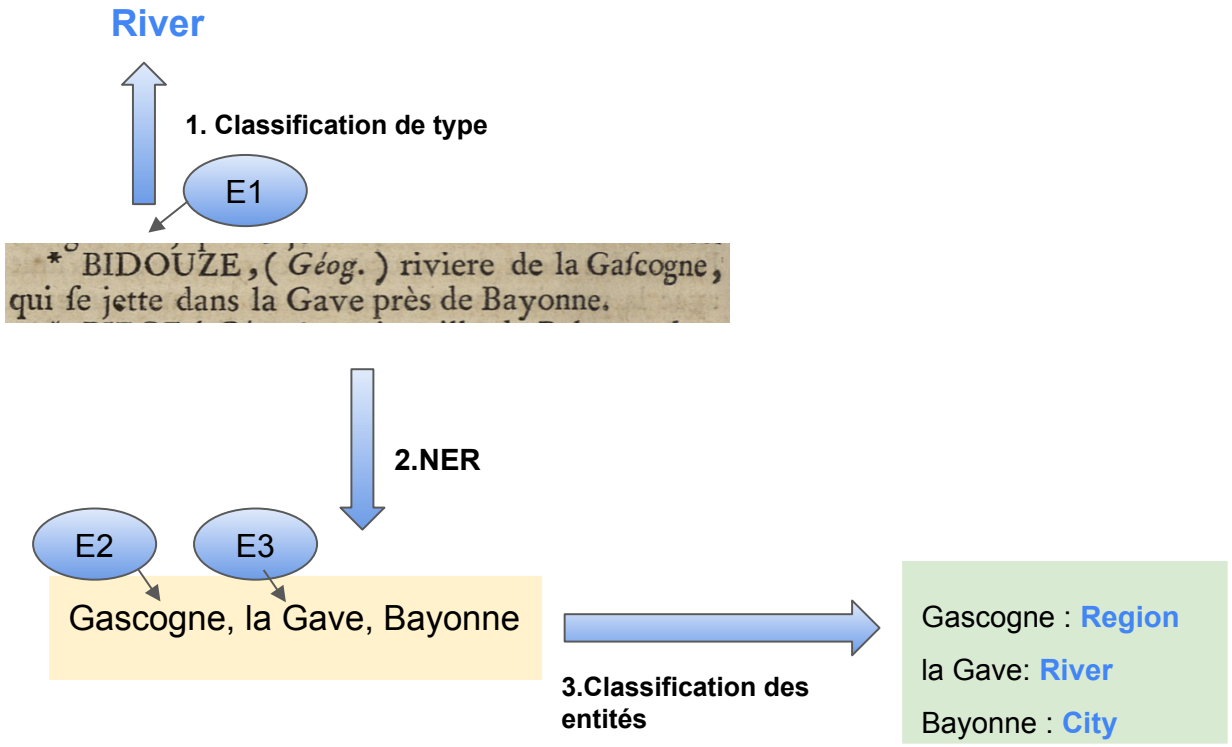


Illustration du workflow d'extraction d'information

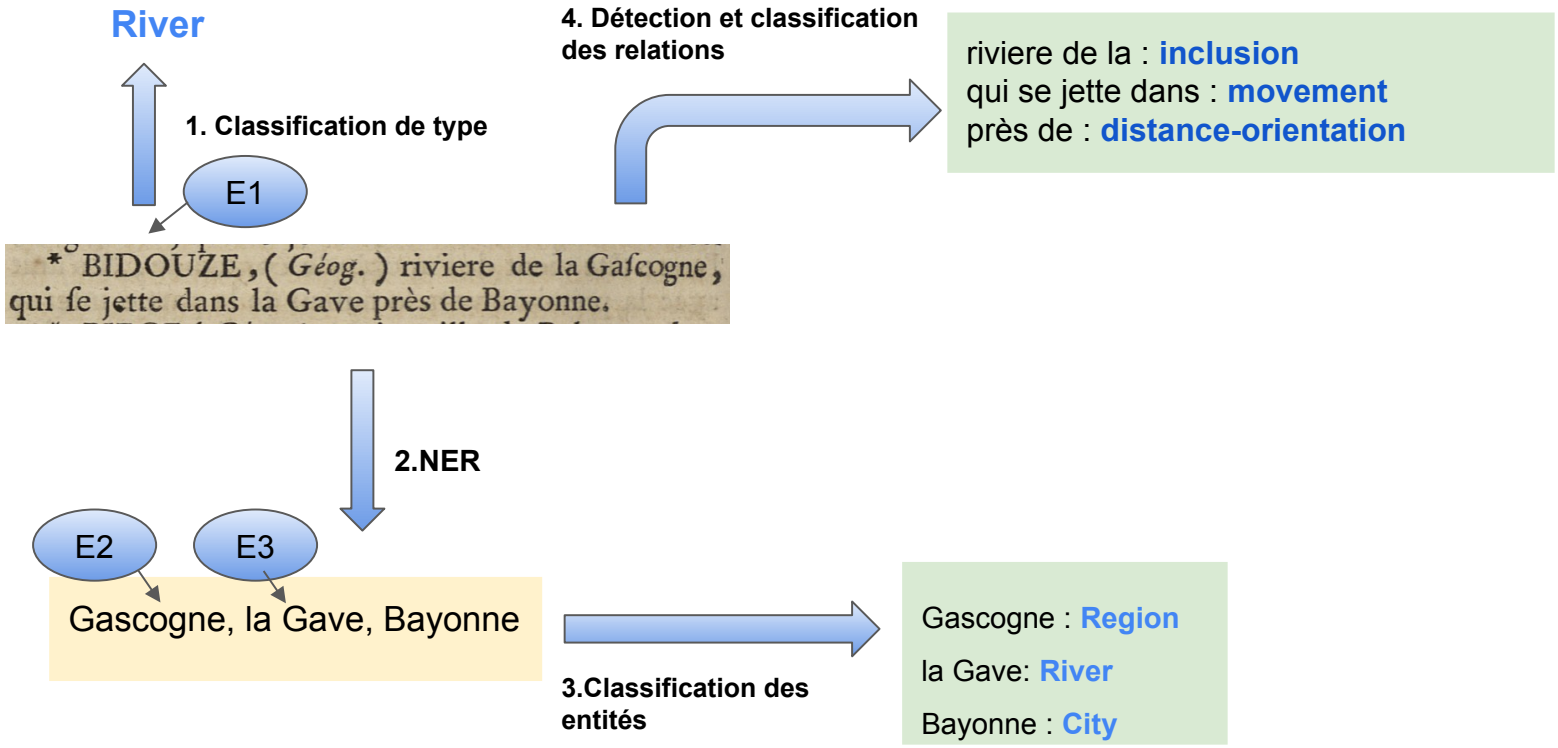
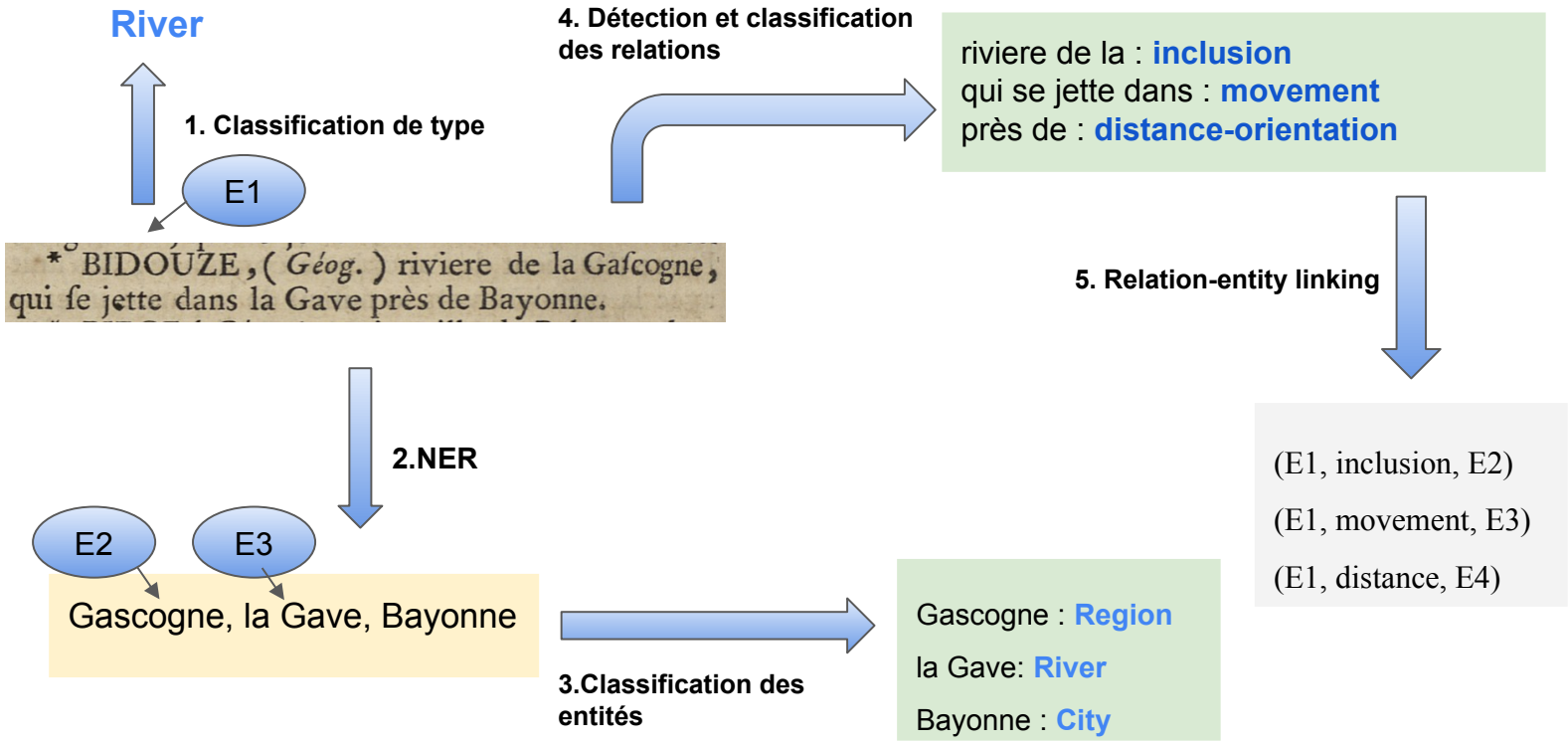
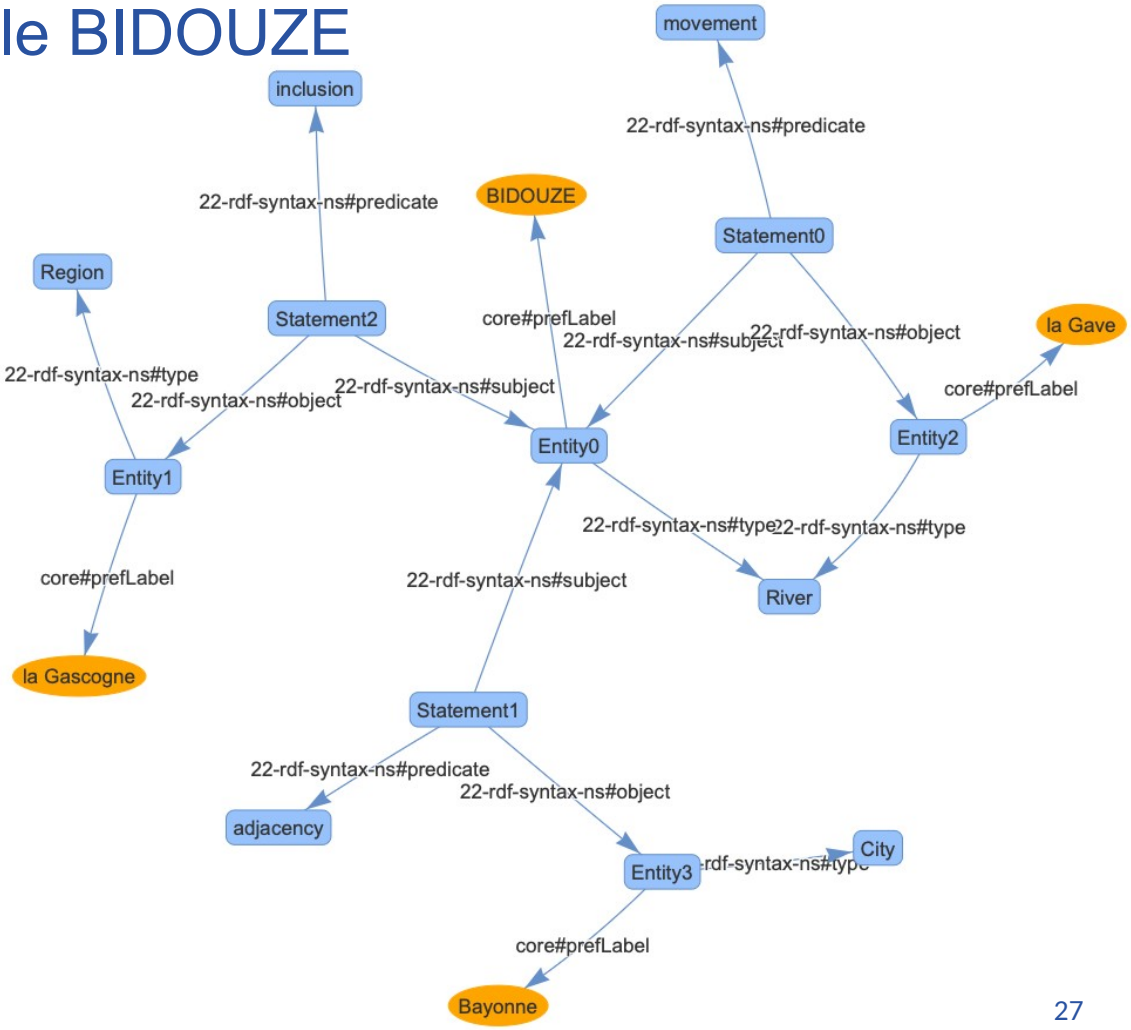


Illustration du workflow d'extraction d'information



Graphe généré pour l'article BIDOUZE

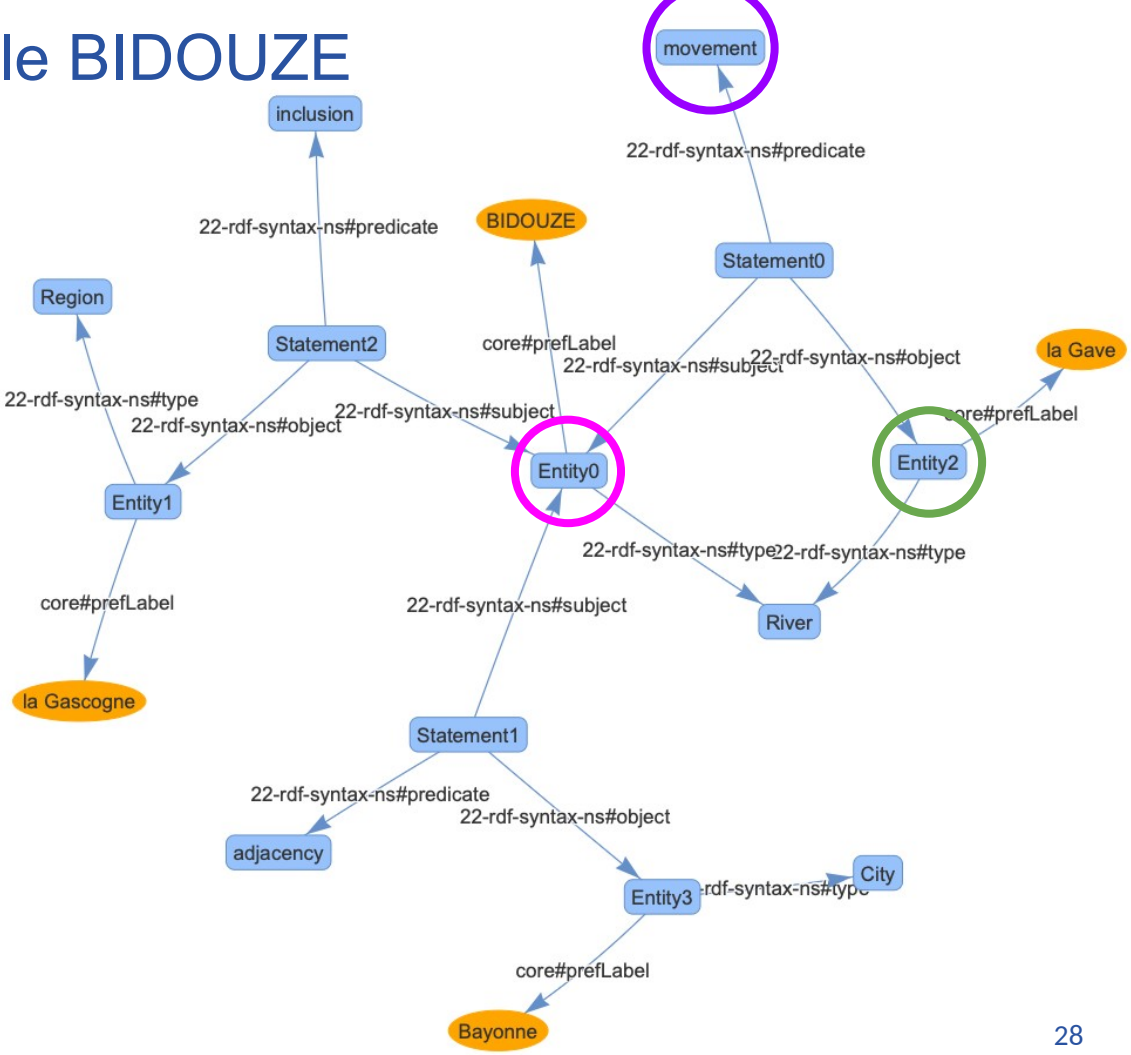
* BIDOUZE, (Géog.) riviere de la Gascogne, qui se jette dans la Gave près de Bayonne.



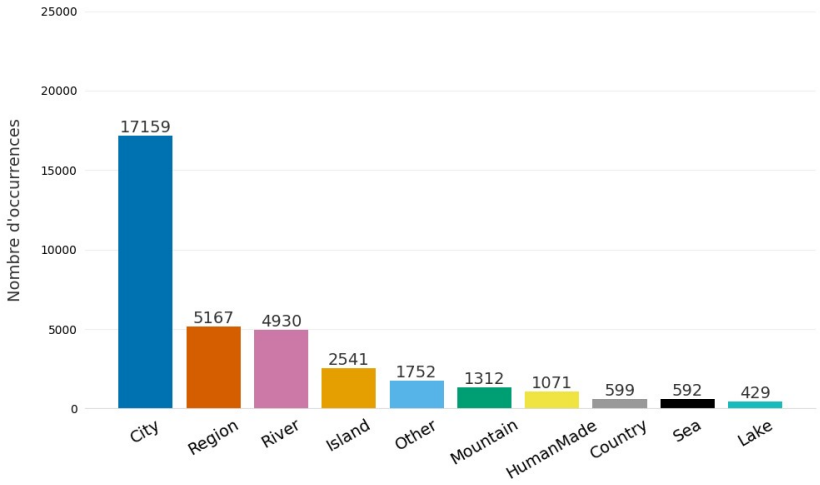
Graphe généré pour l'article BIDOUZE

* BIDOUZE, (Géog.) riviere de la Gascogne, qui se jette dans la Gave près de Bayonne.

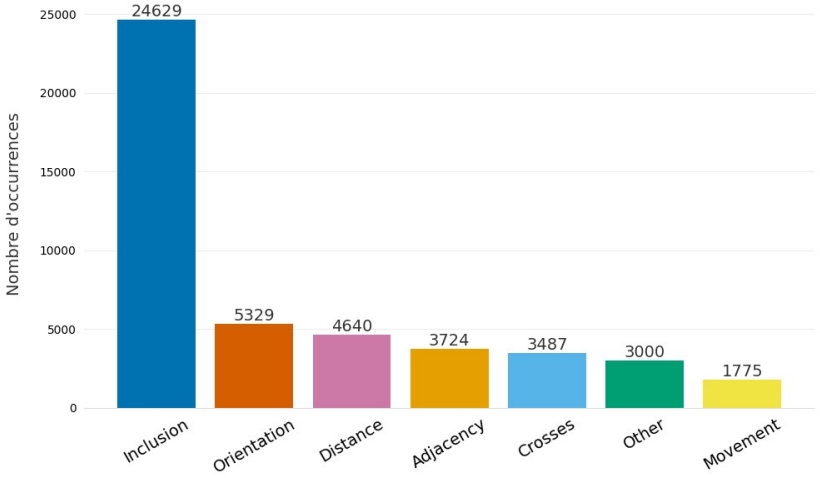
(Entity0, movement, Entity2)



Résultats sur l’ensemble du corpus



Distribution des types d’entités



Distribution des types de relations

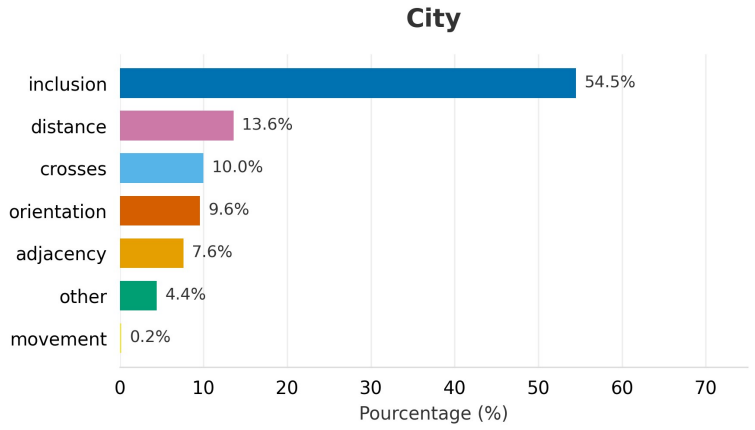
➢ **Entités dominantes:**

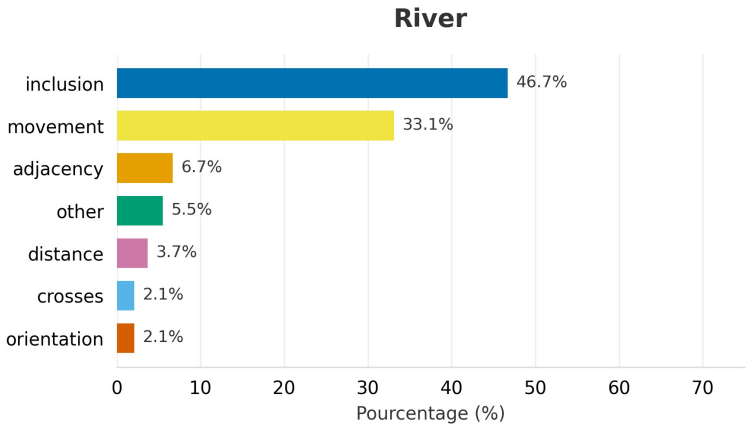
- Ville (48.3%)
- Région (14.5%)
- Rivière(13.9%)

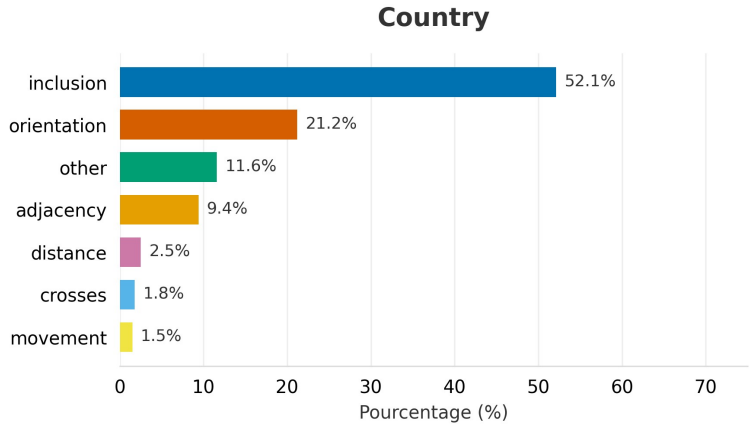
➢ **Relations dominantes:**

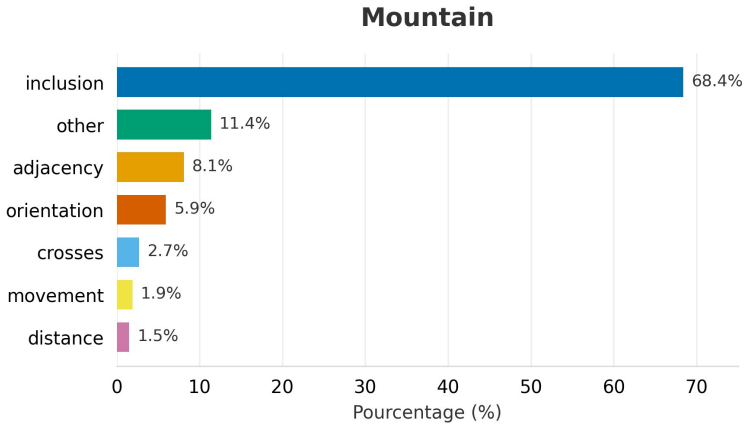
- Inclusion (52.9%)
- Orientation (11.4%)
- Distance (10.0%)
- Adjacence (8.0%)

Répartition des types de relations par type de sujets









Conclusion

Résultats

- Une chaîne de traitement :
  - 8 SLM *fine-tunés* (BERT, T5)
  - un LLM en *few-shot* pour la segmentation d'articles

- Un jeu de données expertisé (différents niveaux de classification)
  - 2750 articles

- Un graphe de connaissances généré :
  - 15 384 articles (domaine Géographie)
  - 35 552 entités
  - 46 584 relations

Perspectives

- Gestion avancée des ambiguïtés
- Amélioration de la modélisation des distances et des unités
- Interopérabilité avec des ontologies spatiales existantes
- Modélisation de l'évolution pour l'analyse diachronique

Démonstration

Extracted informations for entry: Jean-pied-de-Port, S.

- Domain of entry (Géographie / Histoire / Médecine...): **Géographie**
- Type of entry (Place / Person / Other): **Place**
- Type of place (City / River / Mountain / Other...): **City**
- Cardinality (Single / Multiple): **Single**

Named Entity Recognition

Jean-pied-de-Port, S. (Géog.) HEAD ville NC_SPATIAL de France NP_SPATIAL en RELATION Gascogne NP_SPATIAL , à une lieue des RELATION frontieres NC_SPATIAL d' Espagne NP_SPATIAL , autrefois capitale NC_SPATIAL de la basse Navarre NP_SPATIAL , avec une citadelle NC_SPATIAL sur une hauteur. Antonin NP_PERSON appelle ce lieu imus Pyrenoeus NP_SPATIAL , le pié NC_SPATIAL des Pyrénées NP_SPATIAL , parce qu'en eff et il est au pié de cette chaîne de montagnes NC_SPATIAL ; dans ce pays NC_SPATIAL -là on appelle port les passages NC_SPATIAL ou défilés NC_SPATIAL par où l'on peut traverser les Pyrénées NP_SPATIAL , & comme cette ville NC_SPATIAL de S. NC_PERSON Jean NP_PERSON est à l'entrée de ces ports NC_SPATIAL ou passages NC_SPATIAL , on la nomme S. Jean-pied-de port NP_SPATIAL ; elle est sur RELATION la Nive NP_SPATIAL , à l'entrée d RELATION 'un des passages NC_SPATIAL des Pyrénées NP_SPATIAL , à 8 lieues S. E. de RELATION Bayonne NP_SPATIAL , 12 N. E. de RELATION Pampelune NP_SPATIAL , 176 S. O. de RELATION Paris NP_SPATIAL . Long. 16. 22. lat. 43. 8 LATLONG . (D. J.)

Placename	Type
France	Country
Gascogne	Region
Espagne	Country
la basse Navarre	Region
Pyrenoeus	Mountain
Pyrénées	Mountain
les Pyrénées	Mountain
S. Jean-pied-de port	Human-made
la Nive	River
Pyrénées	Mountain
Bayonne	City
Pampelune	City

Relation	Type
en	Inclusion
à une lieue des	Distance-Orientation
sur	Crosses
à l'entrée d	Adjacency
à 8 lieues S. E. de	Distance-Orientation
12 N. E. de	Distance-Orientation
176 S. O. de	Distance-Orientation

RDF graph and map outputs

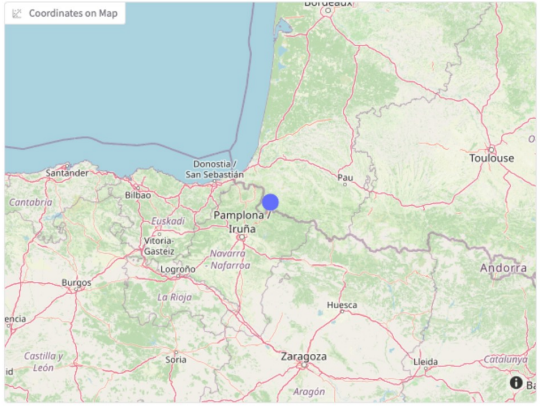
Download Interactive Graph (HTML)

tmpr9kz3bin.html 698.8 KB

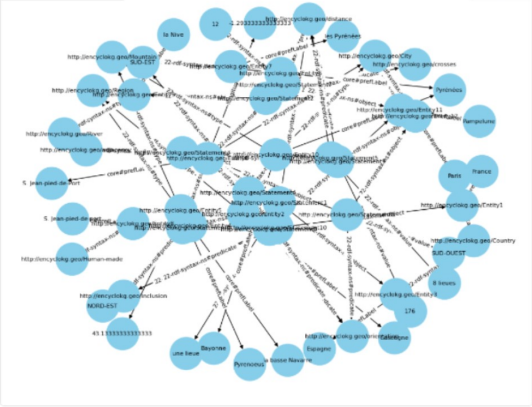
Download RDF file (Turtle)

tmpm7m47zin.ttl 2.5 KB

Coordinates on Map



Static Graph Preview



https://huggingface.co/spaces/GEODE/encyclopedia-to-geokg

Construction automatique d'un graphe de connaissances géo-historiques à partir de textes encyclopédiques anciens

Bin Yang, Ludovic Moncla, Fabien Duchateau, Frédérique Laforest

INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69622 Villeurbanne, France

prenom.nom@liris.cnrs.fr

Modèles & jeux de données

Code

Démonstration

https://huggingface.co/GEODE

https://gitlab.liris.cnrs.fr/ecoda/encyclopedia2geokg

https://huggingface.co/spaces/GEODE/encyclopedia-to-geokg