

Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text

Lucie Cadorel^{1,2,3,4,5}, Alicia Bianchi^{1,3,5,6}, Andrea G. B. Tettamanzi^{1,2,3,4}

Université Nice Côte d'Azur¹, I3S², CNRS³, Inria⁴, KCityLabs⁵, ESPACE⁶

lucie.cadorel@inria.fr



Ingénierie des Connaissances

29 juin 2022, St Etienne, France

Introduction

Contexte



Projet Incertimmo

But : prise en compte de l'incertitude pour une modélisation spatiale des valeurs immobilières dans la ville en utilisant la base de données **Perval**

Limitations : **Perval** est très coûteuse. Problème pour passer à l'échelle nationale

Solution : annonces immobilières publiées sur la Toile mais pas toujours localisées...

<http://imredd.fr/projet/incertimmo/>

Introduction

Contexte

Pourquoi
utiliser les
annonces
immobilières ?



Informations à propos des caractéristiques
du bien



Données exhaustives, à jours et massives



Version en ligne qui facilite la récupération
des données

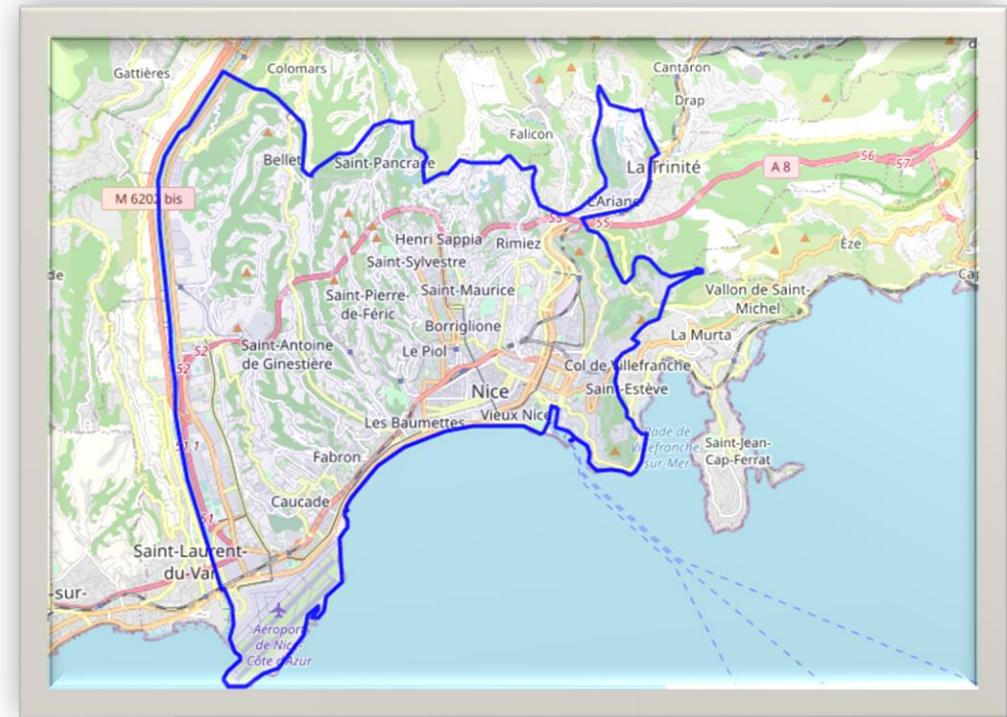


Informations sur la localisation dans le
texte

Introduction

Contexte

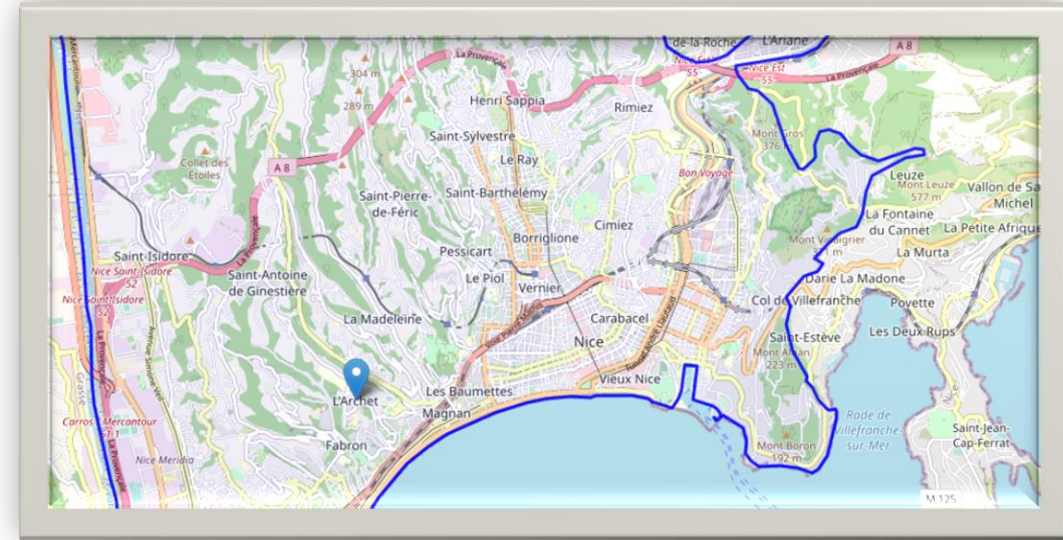
A l'écart de l'agitation citadine, à quelques minutes du coeur de la ville de Nice, venez découvrir cet appartement situé sur L'Archet, l'une des plus belles collines niçoises. Cet appartement de 64m² est ouvert sur une belle terrasse [...]



Introduction

Contexte

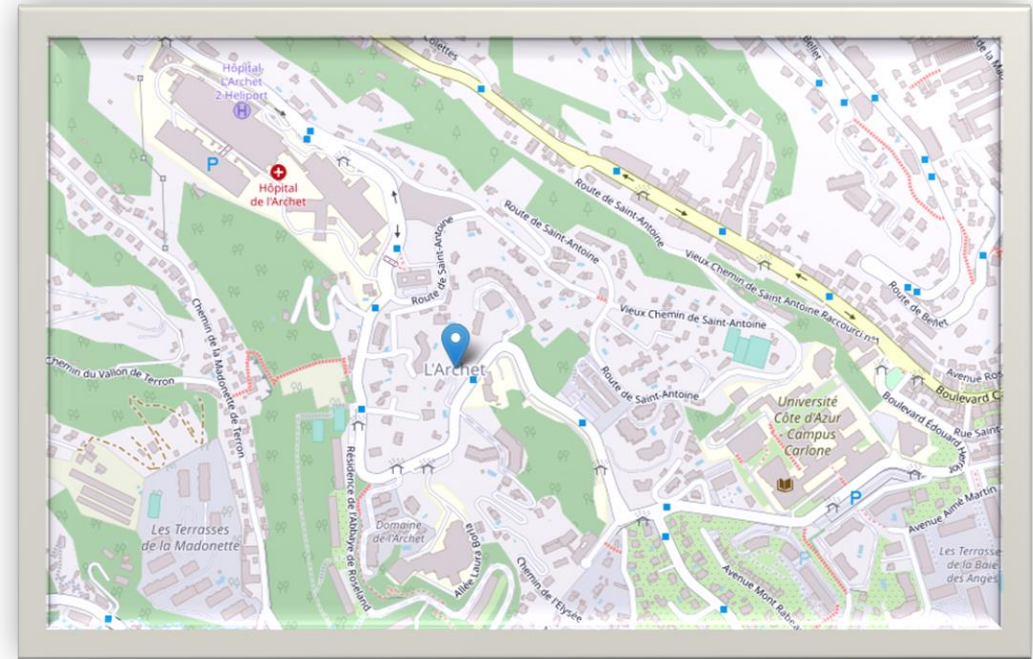
A l'écart de l'agitation citadine, à quelques minutes du coeur de la ville de Nice, venez découvrir cet appartement situé sur L'Archet, l'une des plus belles collines niçoises. Cet appartement de 64m² est ouvert sur une belle terrasse [...]



Introduction

Contexte

A l'écart de l'agitation citadine, à quelques minutes du coeur de la ville de Nice, venez découvrir cet appartement situé sur L'Archet, l'une des plus belles collines niçoises. Cet appartement de 64m² est ouvert sur une belle terrasse [...]



Introduction

Background

Named Entity Recognition

- Patrons linguistiques, Dictionnaires, Deep Learning
- Catégories : Personne, Toponyme, Organisation,...

Texte court et en français

- Avancées avec modèles de langages (*e.g.*, CamemBERT)
- Domaines : Legal text (*Barrière et Fouret, 2019*), guide de randonnées (*Moncla et al., 2014*), etc.

Entités spatiales

- En général, toponyme
- Absolue vs Relative (*Lesbegueries et al., 2006*)
- Extended Named Entity (*Gaio et Moncla, 2017*)

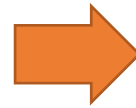
Méthodologie



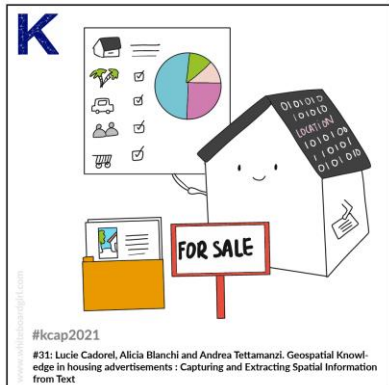
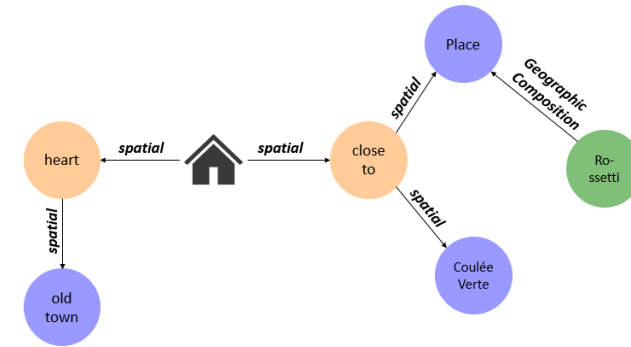
Annotated housing ads in French



Named Entity Recognition



Relationship Extraction



Publication : Lucie Cadorel, Alicia Bianchi, and Andrea G. B. Tettamanzi. 2021. Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. In Proceedings of the 11th on Knowledge Capture Conference (K-CAP '21). Association for Computing Machinery, New York, NY, USA, 41–48.
<https://doi.org/10.1145/3460210.3493547>

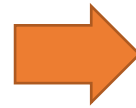
Méthodologie



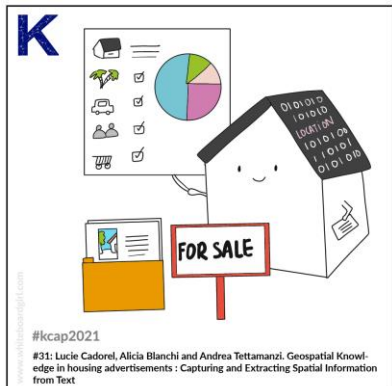
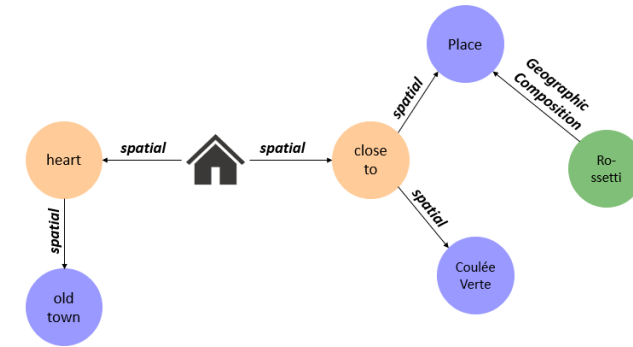
Annotated housing ads in French



Named Entity Recognition



Relationship Extraction



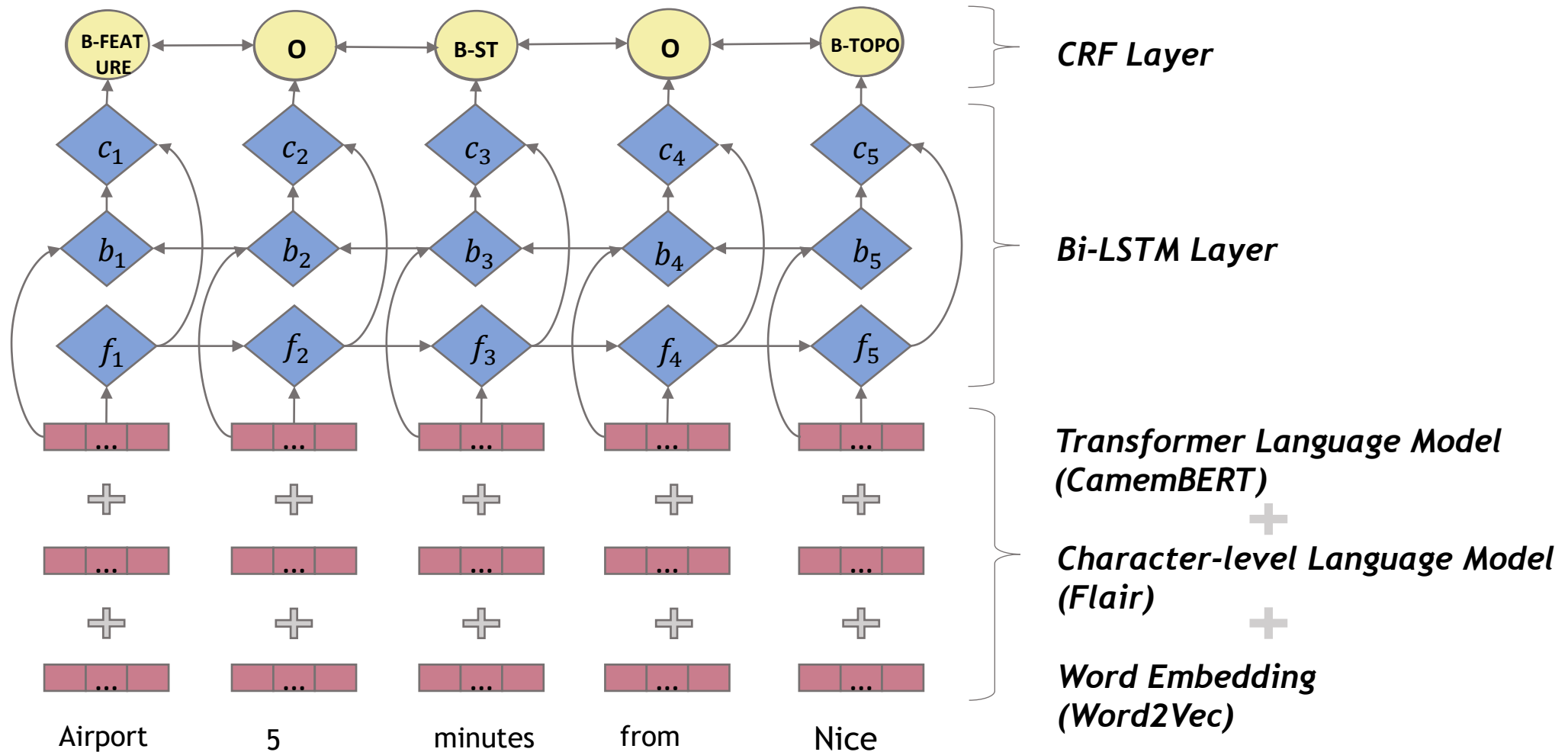
Publication : Lucie Cadorel, Alicia Bianchi, and Andrea G. B. Tettamanzi. 2021. Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. In Proceedings of the 11th on Knowledge Capture Conference (K-CAP '21). Association for Computing Machinery, New York, NY, USA, 41–48.
<https://doi.org/10.1145/3460210.3493547>

Entités géographiques et spatiales

- **Toponym** : Noms propres des lieux
- **Feature** : Type de lieux (nature, constructions, quartiers, etc.)
- **Spatio-temporal** : terme désignant la relation spatiale entre deux lieux
- **Mode of transportation** : mode de transport entre deux lieux (à pied, voiture, etc.)

NICE - Appartement proche de la célèbre Promenade des Anglais, et à deux pas du tram. Les commerces et écoles sont à 5 minutes à pied.

Modèle de NER



Résultats

	Model	Precision	Recall	F1-Score
Spacy	Pre-trained French model	0.830 (0.03)	0.822 (0.03)	0.821 (0.02)
	Own training	0.828 (0.02)	0.845 (0.01)	0.835 (0.01)
Bi-LSTM - CRF	Word2Vec	0.786 (0.02)	0.741 (0.02)	0.763 (0.02)
	Flair	0.833 (0.01)	0.876 (0.01)	0.854 (0.01)
	CamemBERT	0.851 (0.01)	0.877 (0.005)	0.865 (0.01)
	Flair + Word2Vec	0.837 (0.01)	0.864 (0.01)	0.85 (0.01)
	Camembert + Word2Vec	0.860 (0.01)	0.872 (0.004)	0.866 (0.005)
	Flair + CamemBERT	0.861 (0.01)	0.884 (0.02)	0.872 (0.01)
	Flair + CamemBERT + Word2Vec	0.863 (0.005)	0.889 (0.01)	0.876 (0.01)

Table 3: Performance of NER models.

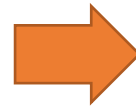
Méthodologie



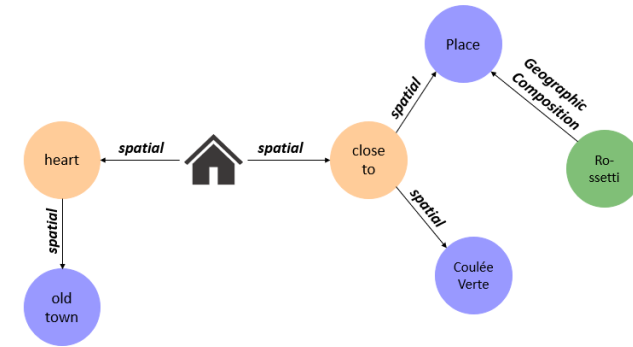
Annotated housing ads in French



Named Entity Recognition



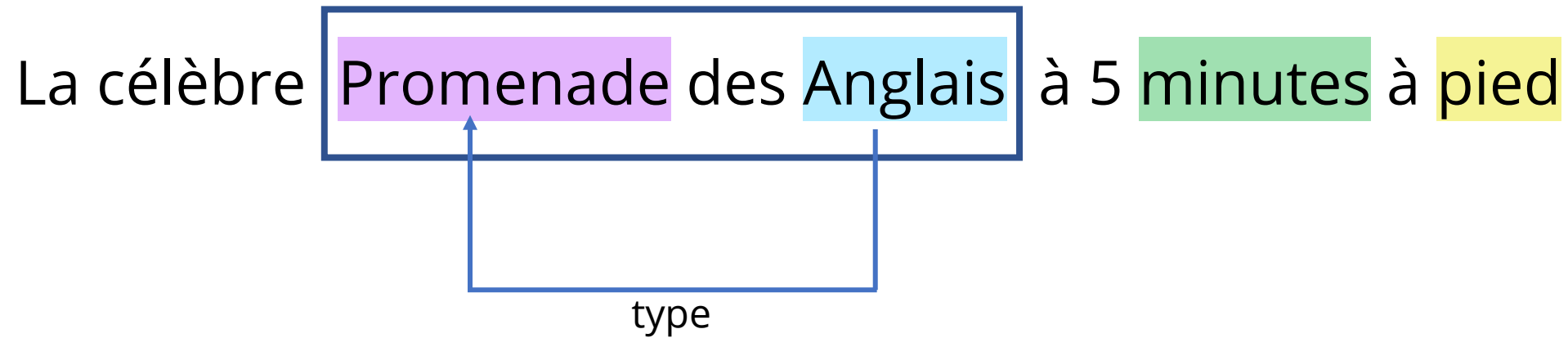
Relationship Extraction



Publication : *Lucie Cadorel, Alicia Bianchi, and Andrea G. B. Tettamanzi. 2021. Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. In Proceedings of the 11th on Knowledge Capture Conference (K-CAP '21). Association for Computing Machinery, New York, NY, USA, 41–48. <https://doi.org/10.1145/3460210.3493547>*

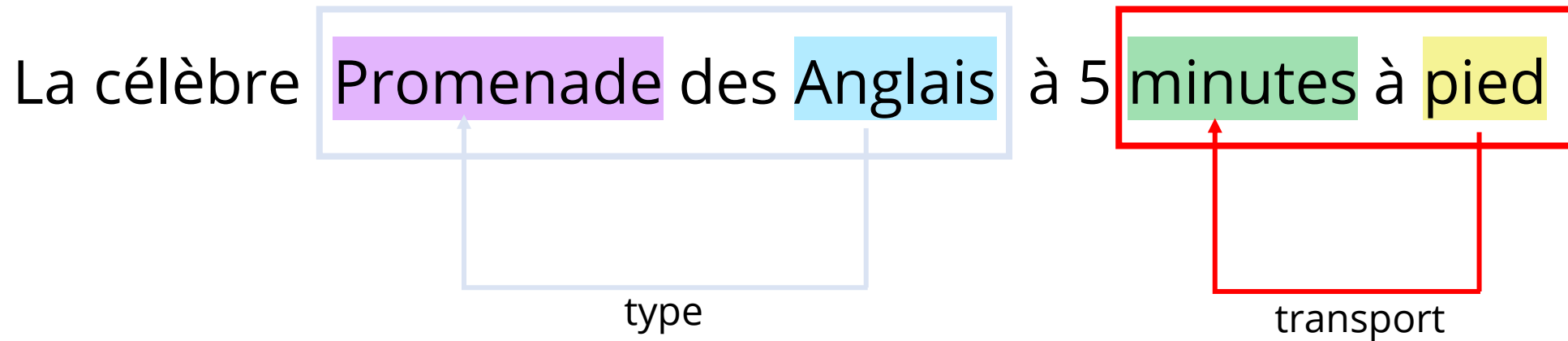
Extraction de Relations

Types de relations



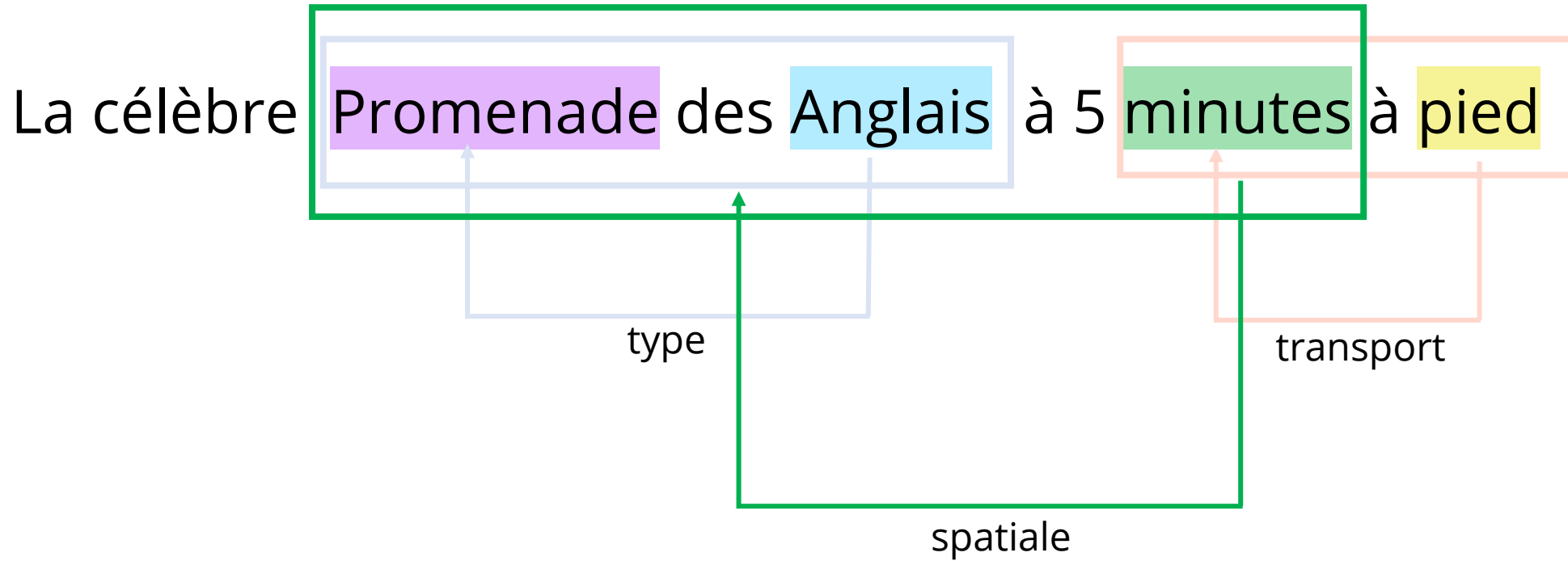
Extraction de Relations

Types de relations



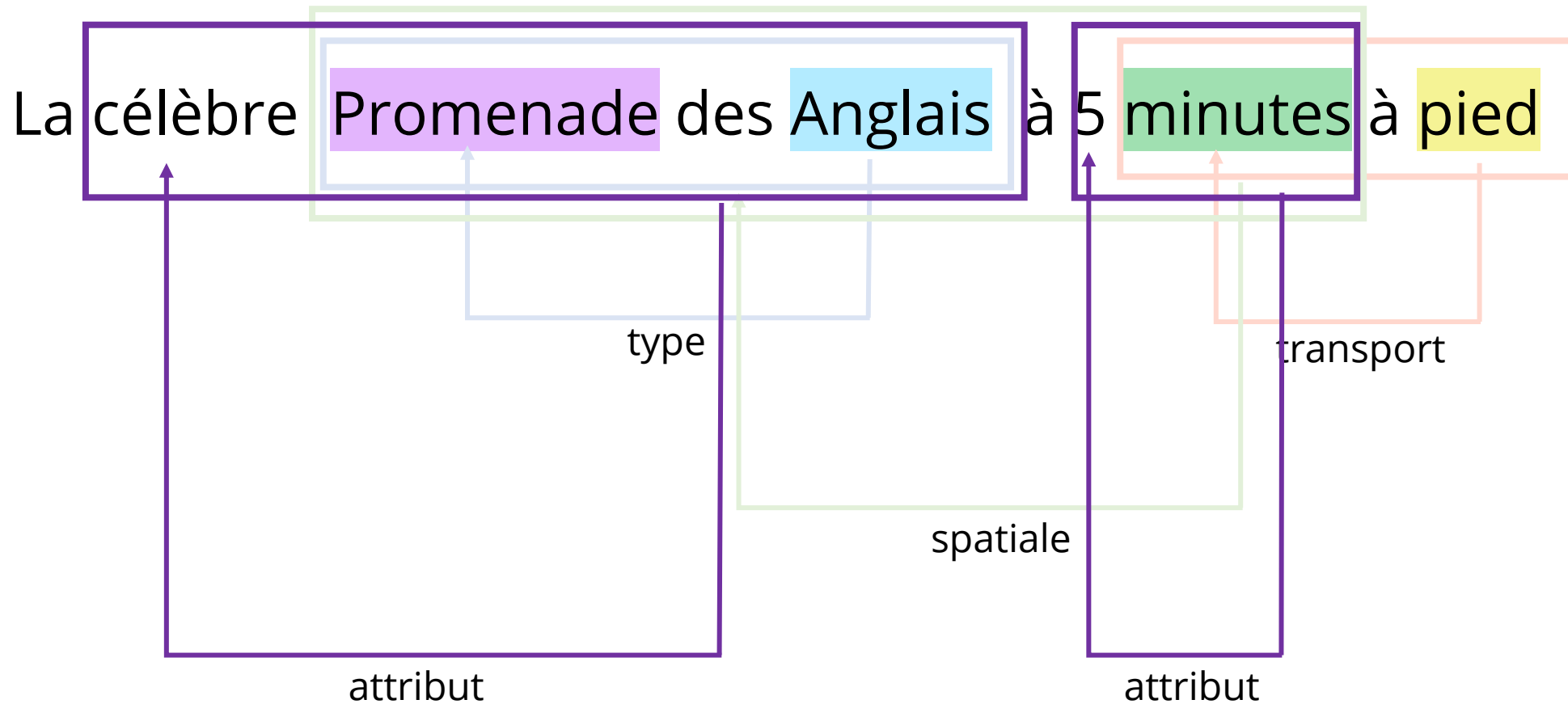
Extraction de Relations

Types de relations



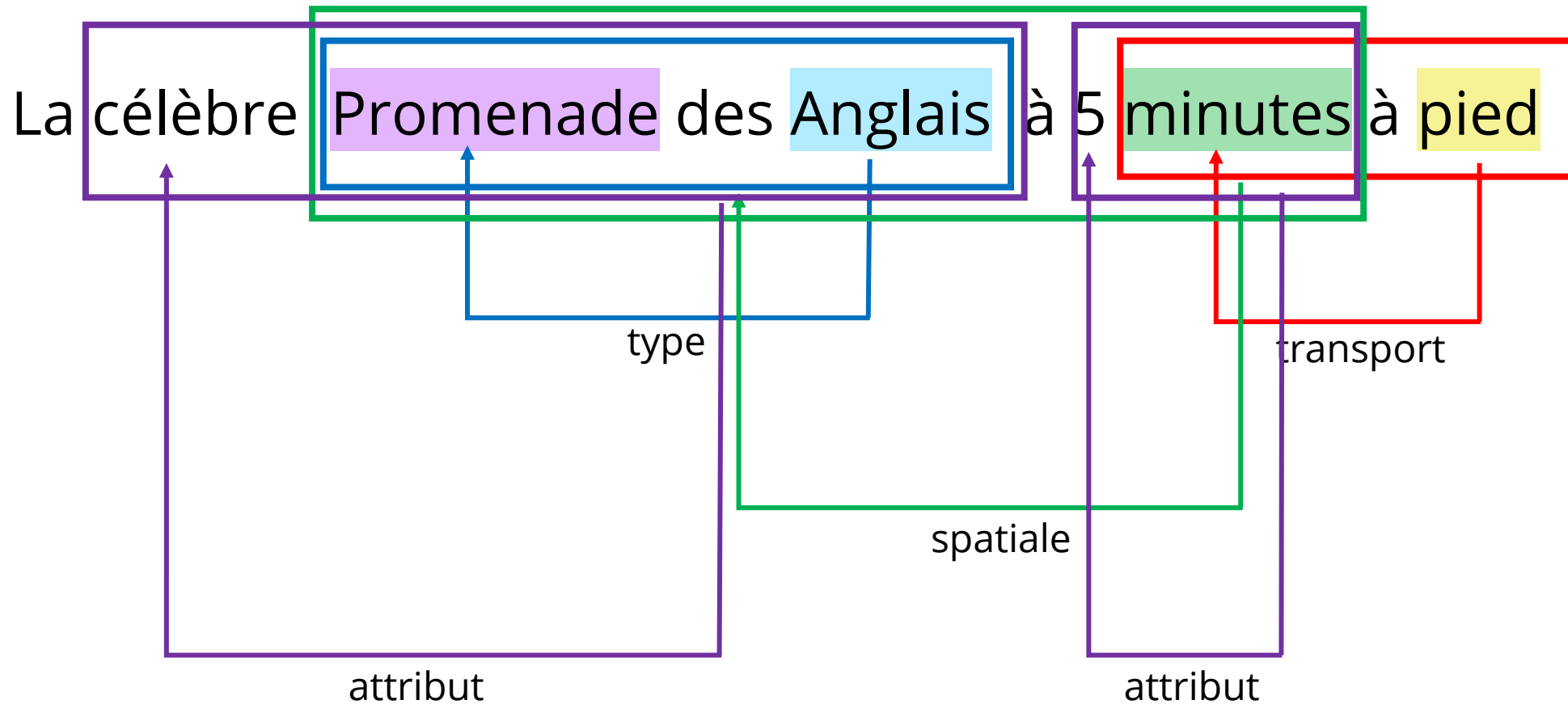
Extraction de Relations

Types de relations



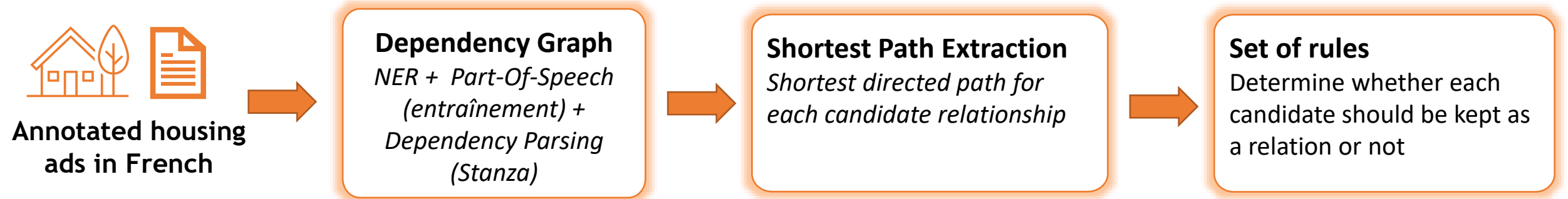
Extraction de Relations

Types de relations

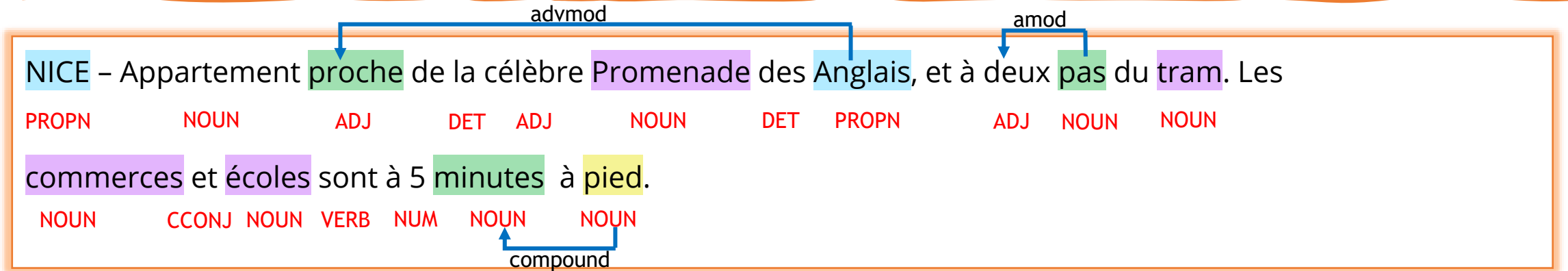


Extraction de Relations

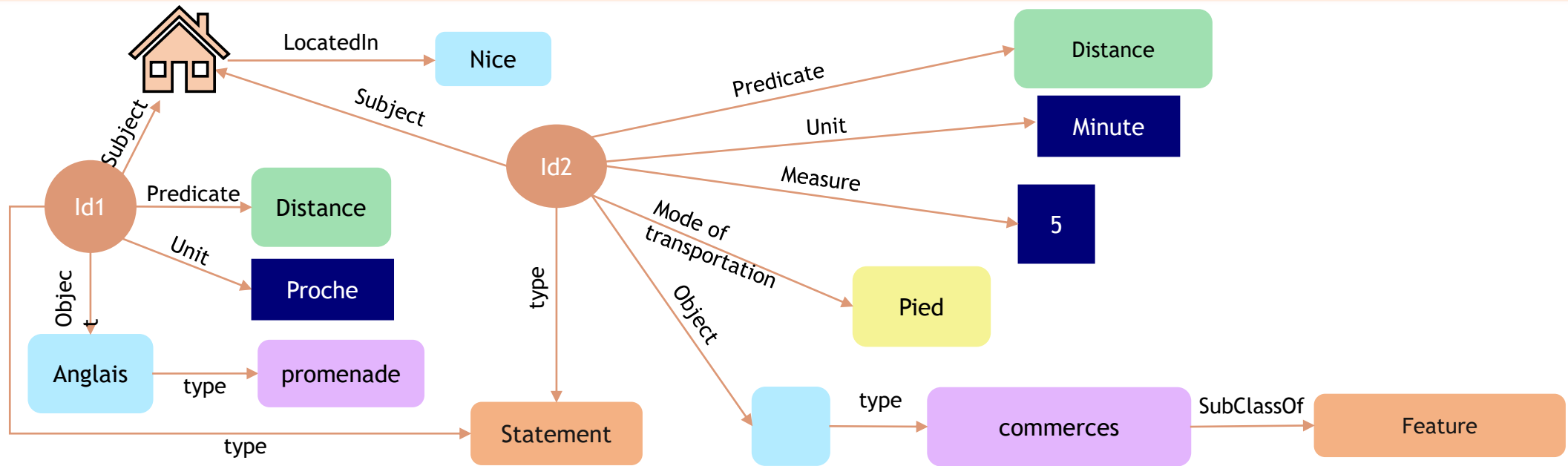
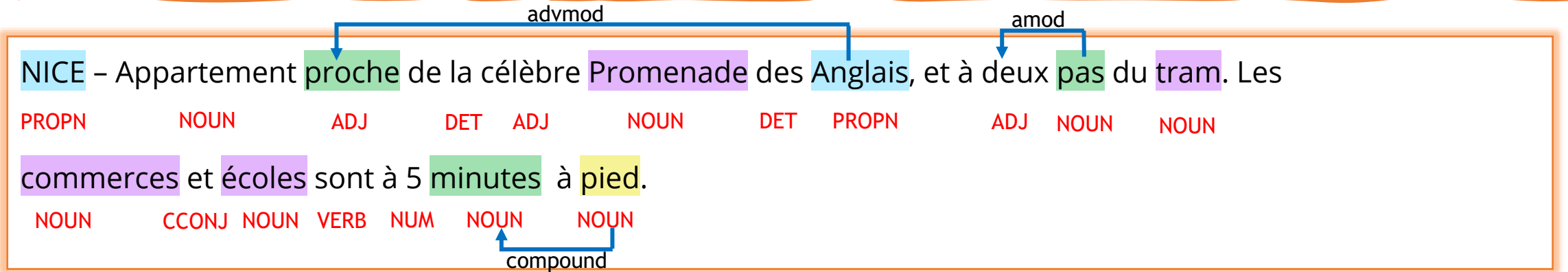
Méthode



Du texte au graphe



Du texte au graphe



Conclusion

Extraction de connaissances spatiales

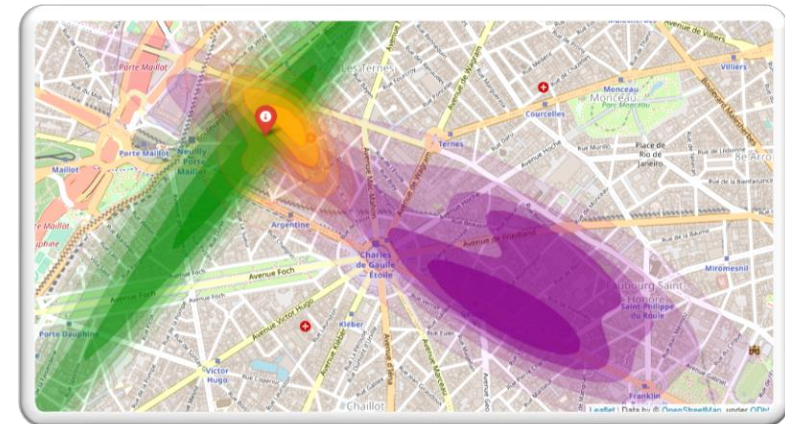
- Named Entity Recognition pour des textes en français
- Définition d'entités spatiales
- Extraction de relations pour structurer les informations

Démo

<http://erebe-vm2.i3s.unice.fr:5000/>

Future Work

- Désambiguation / Résolution des entités spatiales
- Localisation des annonces (en cours...)
- Représentation « floue » des objets spatiaux





Merci pour votre attention !

Lucie Cadorel
lucie.cadorel@inria.fr

Ingénierie des Connaissances
29 juin 2022, St Etienne, France