



Université de Lille

Soutenance de  
mémoire

Génération automatique  
d'exercices de  
syntaxe à partir de  
corpus annotés.

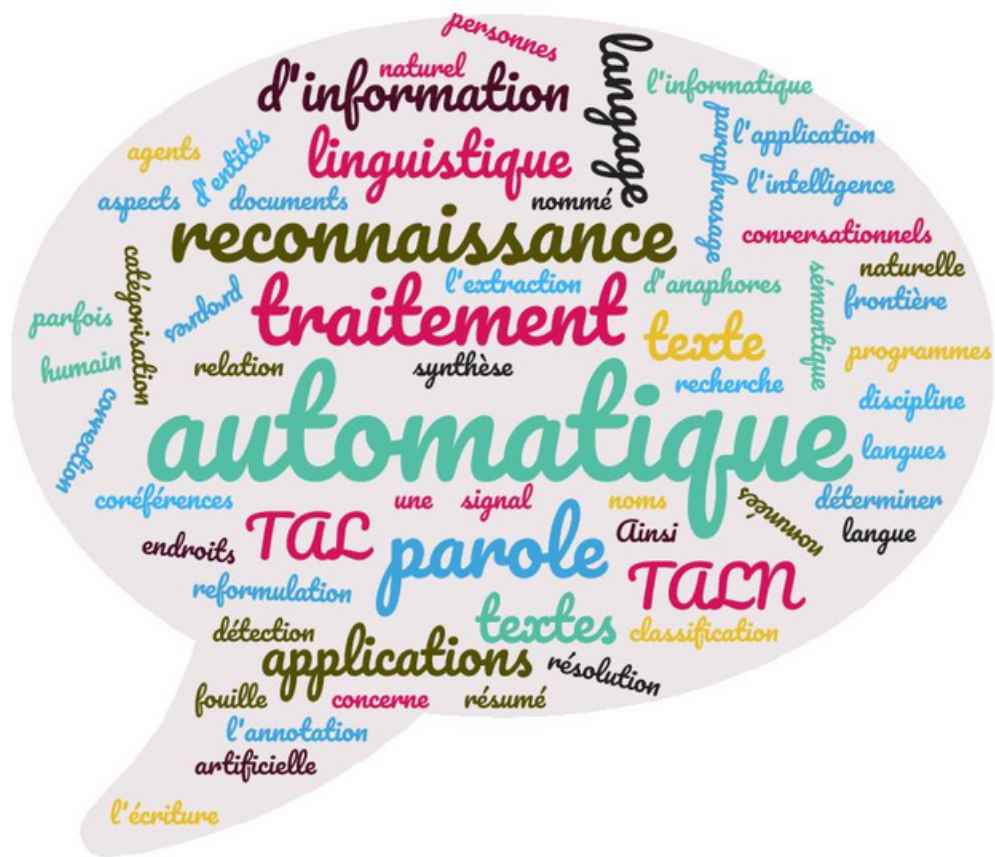
**Pham Thi Thuy**

Master 2 mention SDL parcours LTTAC

**Antonio Balvet:** Maître de conférences et Enseignant Chercheur en Sciences du Langage à l'Université de Lille. (directeur)

**Katia Paykin Arroues:** Maîtresse de conférences et Enseignant Chercheur en Sciences du Langage à l'Université de Lille. (examinatrice)

# PLAN



# 1. Introduction

## 2. Problématique

### 3. Etat de l'art: H5P, Moodle, WIMS

## 4. Évaluation des corpus avec des corpus UDpipe, Spacy, Stanza

## 5. Réalisation technique

## 6. Conclusion

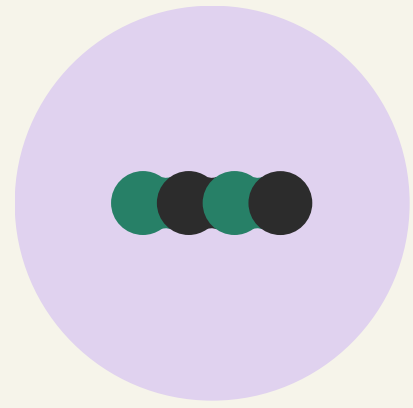
## 7. Questions et réponses

[Retour à la vue d'ensemble](#)



# Introduction

- Pourquoi j'ai choisi ce sujet de mémoire ?



# La problématique

Comment générer automatiquement des exercices de syntaxe à partir de corpus donnée?

[Retour à la vue d'ensemble](#)

## Objectifs

1. Préparation des données à partir d'un corpus pour la création automatisée de Quiz.
2. Création d'une application de WEB permettant de résoudre les inconvénients de plate-form Moodle.

[Retour à la vue d'ensemble](#)

# Etat de l'art



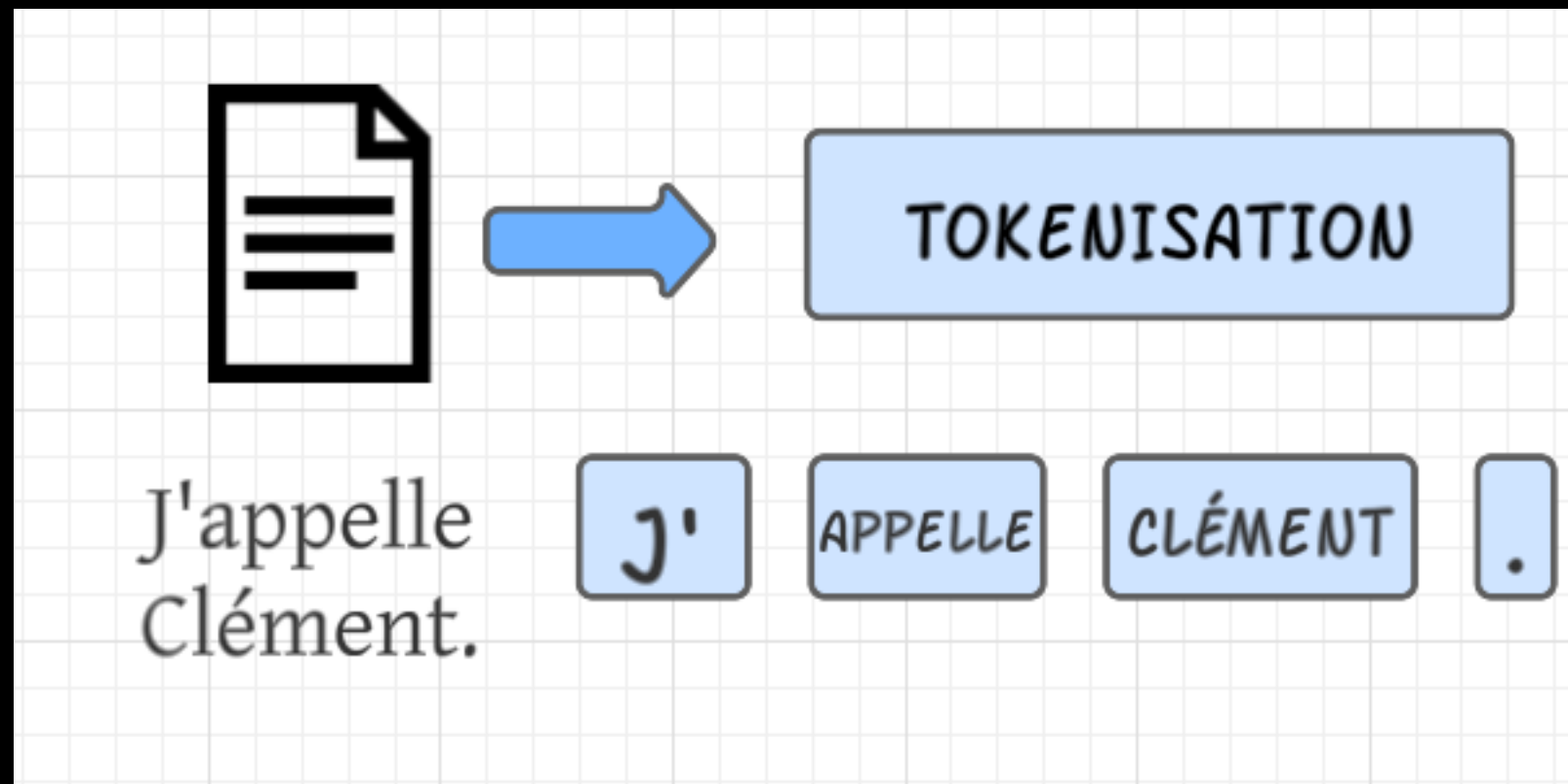
H5P, Moodle et WIMS sont des plateformes gratuites.

Avantage de H5P, Moodle et WIMS	Inconvénient de H5P, Moodle et WIMS
<ol style="list-style-type: none"><li>1. Open-source</li><li>2. Facile à utiliser</li><li>3. Intégration avec d'autres outils (H5P)</li></ol>	<ol style="list-style-type: none"><li>1. Création des exercices est réalisée manuellement.</li><li>2. Absence de fonctionnalités d'accessibilité.</li><li>3. Impossibilité d'insérer un arbre syntaxique.</li><li>4. Impossibilité d'enregistrer le temps de réponse et le nombre de tentatives pour chaque question.</li></ol>

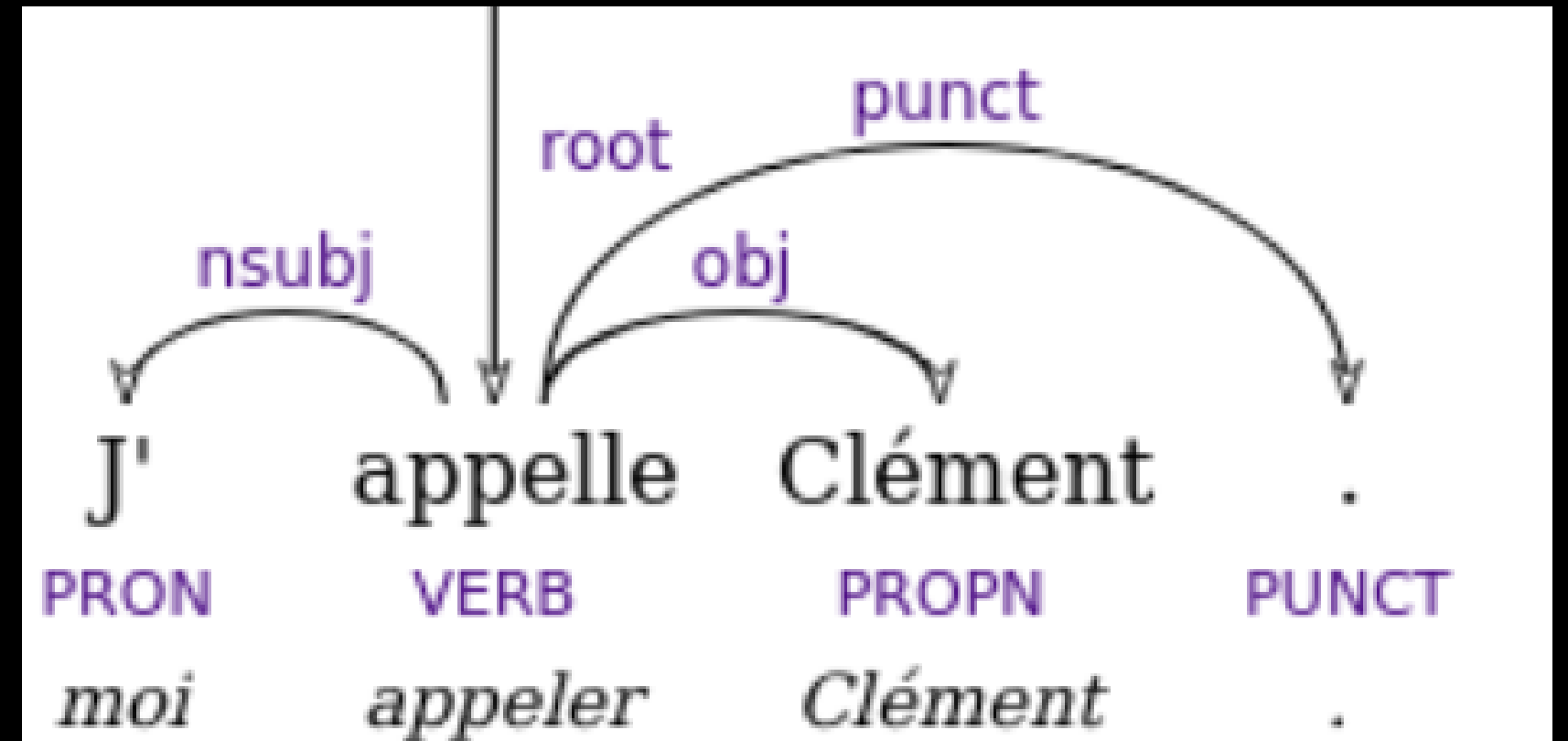
# Universal Dependencies

- un projet collaboratif
- Élaborer un schéma d'annotation syntaxique pour plus de 100 langues différents.

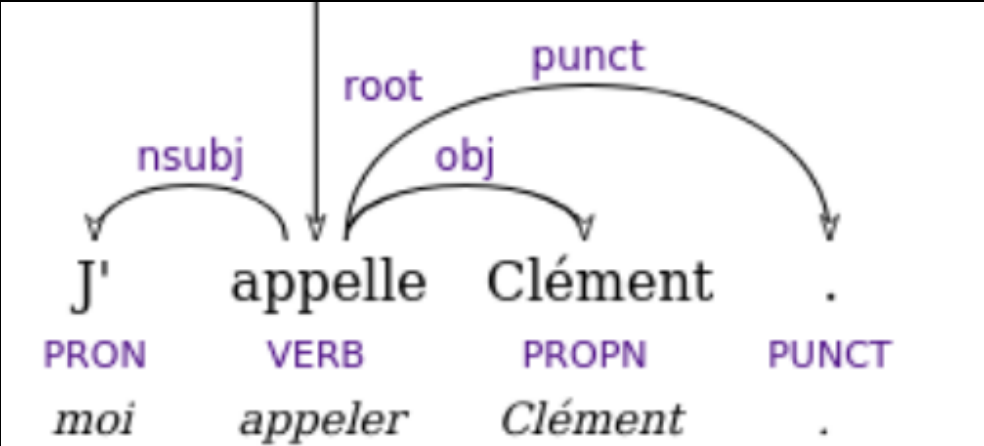
# Segmentation (tokenisation)



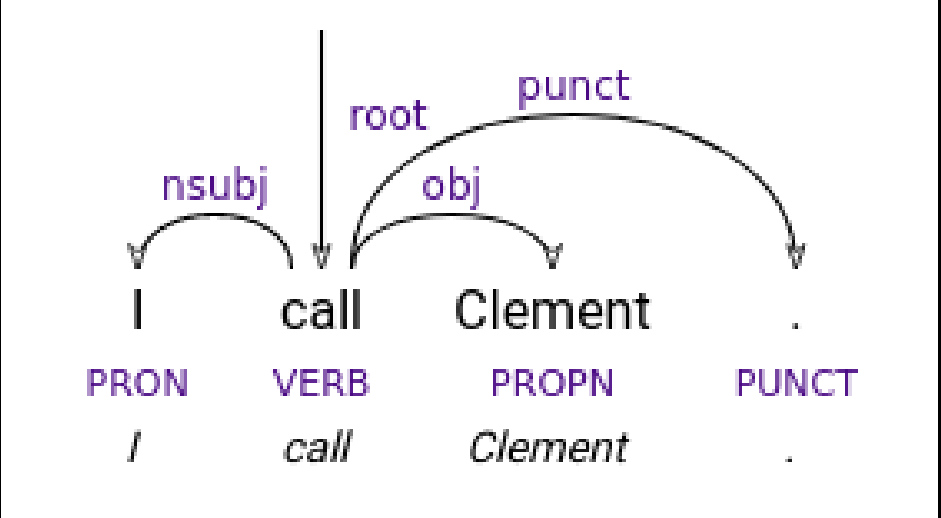
# Dépendance syntaxique



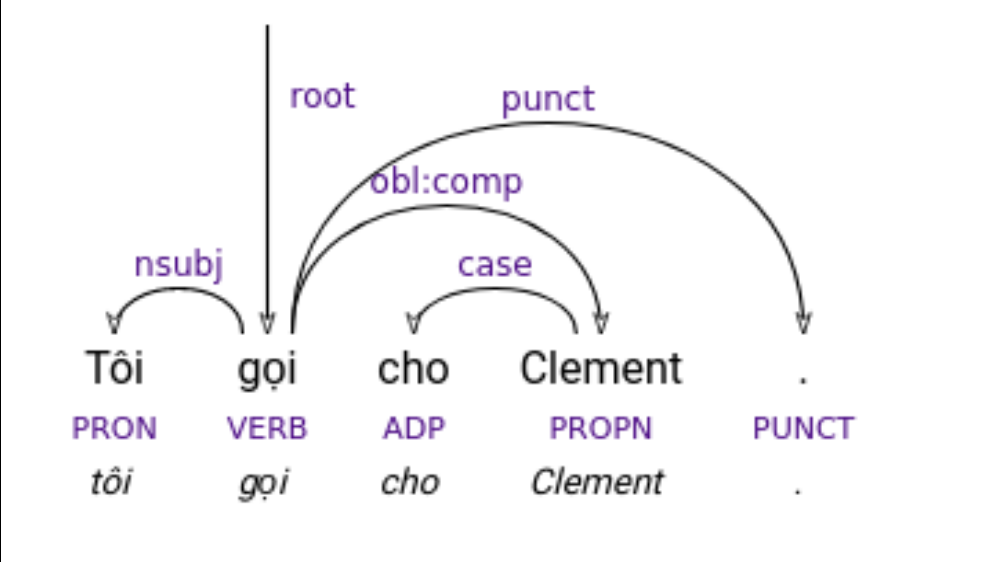
1	J'	moi	PRON	—	Emph=No Number=Sing Person=1 PronType=Prs	2	nsubj
2	appelle	appeler	VERB	—	Mood=Ind Number=Sing Person=1 Tense=Pres VerbForm=Fin	0	root
3	Clément	Clément	PROPN	—	—	2	obj
4	.	.	PUNCT	—	—	2	punct



1	I	I	PRON	PRP	Case=Nom Number=Sing Person=1 PronType=Prs	2	nsubj
2	call	call	VERB	VBP	Mood=Ind Number=Sing Person=1 Tense=Pres VerbForm=Fin	0	root
3	Clement	Clement	PROPN	NNP	Number=Sing	2	obj
4	.	.	PUNCT	.	—	2	punct



1	Tôi	tôi	PRON	Pro	—	2	nsubj
2	gọi	gọi	VERB	V	—	0	root
3	cho	cho	ADP	Pre	—	4	case
4	Clement	Clement	PROPN	NNP	—	2	obl:comp
5	.	.	PUNCT	.	—	2	punct





# Outils

**Stanza**

**Stanza est une  
solution de référence.**

**UDPipe**

**UDPipe est une solution  
de référence.**

**Spacy**

**Spacy est une solution  
« industrielle »**

# Comparaison la performance entre : Stanza, UDPipe, Spacy

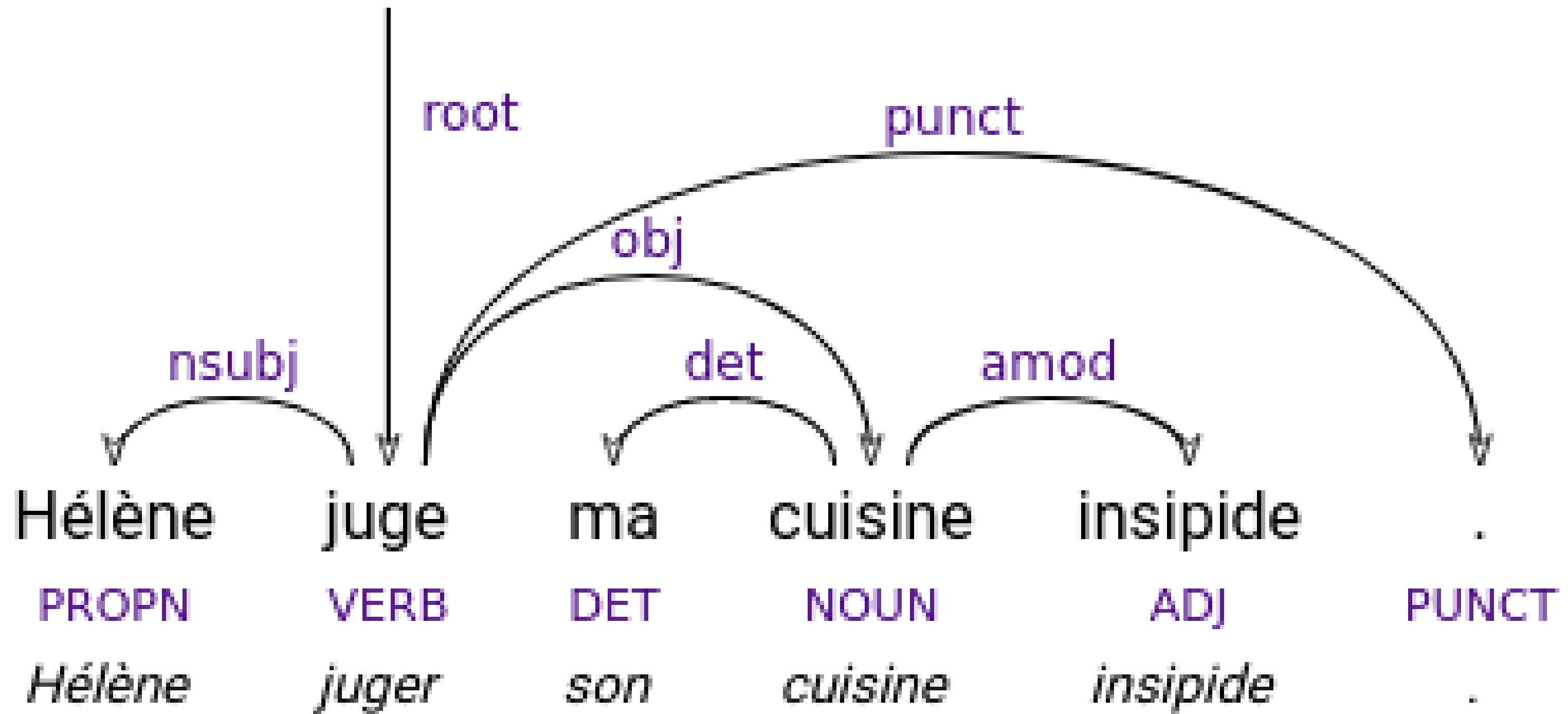
Treebank	System	Tokens	UPOS
French-GSD	Stanza	99.68	97.30
	UDPipe	99.68	95.85
	spaCy	98.34	86.82
English-EWT	Stanza	99.01	95.40
	UDPipe	98.90	93.26
	spaCy	97.30	86.72

- Évaluation deux corpus: attribut du COD et complément du COD et COI
- Évaluation trois outils : UDPipe, Spacy, Stanza

# Attribut du COD : UD Pipe

# text = Hélène juge ma cuisine insipide.										
1	Hélène	Hélène	PROPN	_	_		2	nsubj	_	TokenRange=0:6
2	juge	juger	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin		0	root	_	TokenRange=7:11
3	ma	son	DET	_	Gender=Fem Number=Sing Number[psor]=Sing Person[psor]=1 Poss=Yes PronType=Prs		4	det	_	TokenRange=12:14
4	cuisine	cuisine	NOUN	_	Gender=Fem Number=Sing		2	obj	_	TokenRange=15:22
5	insipide	insipide	ADJ	_	Gender=Fem Number=Sing		4	amod	_	SpaceAfter=No TokenRange=23:31
6	.	.	PUNCT	_	_		2	punct	_	SpacesAfter='\r\n' TokenRange=31:32

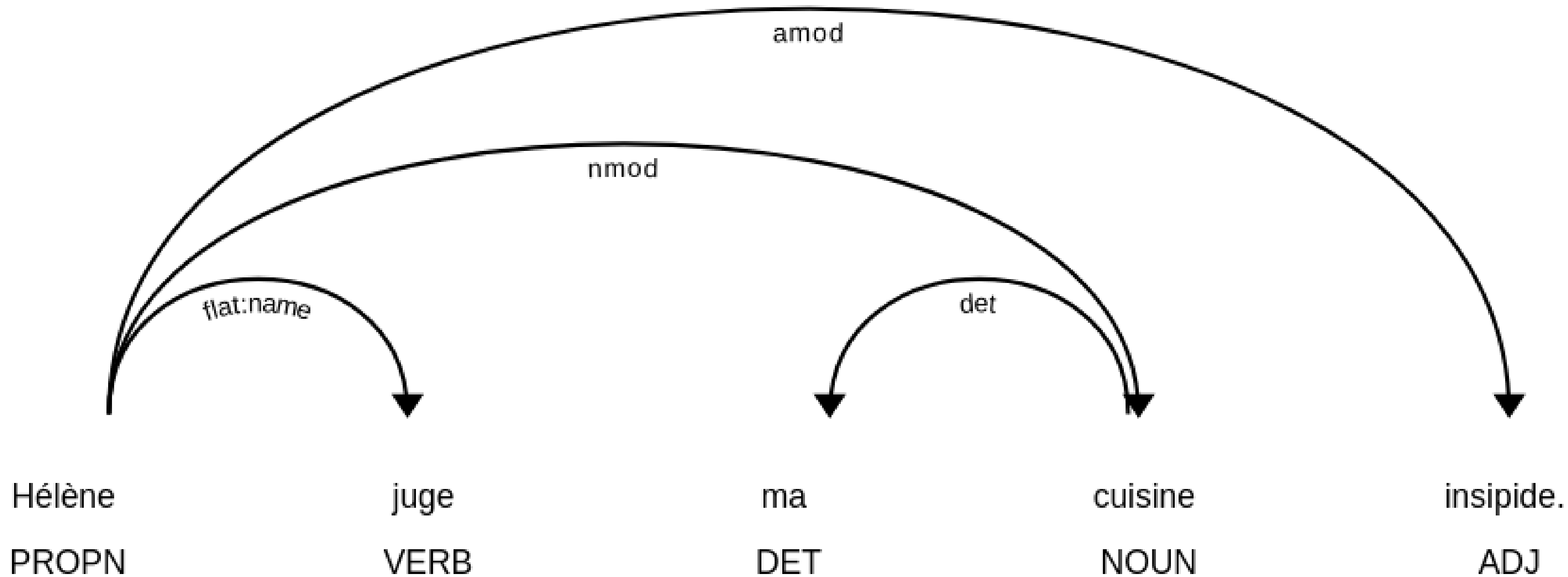
# Attribut du COD : UDPipe



# Attribut du COD : Spacy

ID: 0	TOKEN: Hélène	TAG: proper noun	Head: 0	DEPREL: root
ID: 1	TOKEN: juge	TAG: verb	Head: 0	DEPREL: None
ID: 2	TOKEN: ma	TAG: determiner	Head: 3	DEPREL: determiner
ID: 3	TOKEN: cuisine	TAG: noun	Head: 0	DEPREL: modifier of nominal
ID: 4	TOKEN: insipide	TAG: adjective	Head: 0	DEPREL: adjectival modifier
ID: 5	TOKEN: .	TAG: punctuation	Head: 0	DEPREL: punctuation

# Attribut du COD : Spacy



# Attribut du COD : Stanza

[Sentence 1]

ID: 1    TOKEN: Hélène    LEMMA: Hélène    POS: PROPN    HEAD: 2    DEPREL: nsubj

ID: 2    TOKEN: juge    LEMMA: juger    POS: VERB    HEAD: 0    DEPREL: root

ID: 3    TOKEN: ma    LEMMA: son    POS: DET    HEAD: 4    DEPREL: det

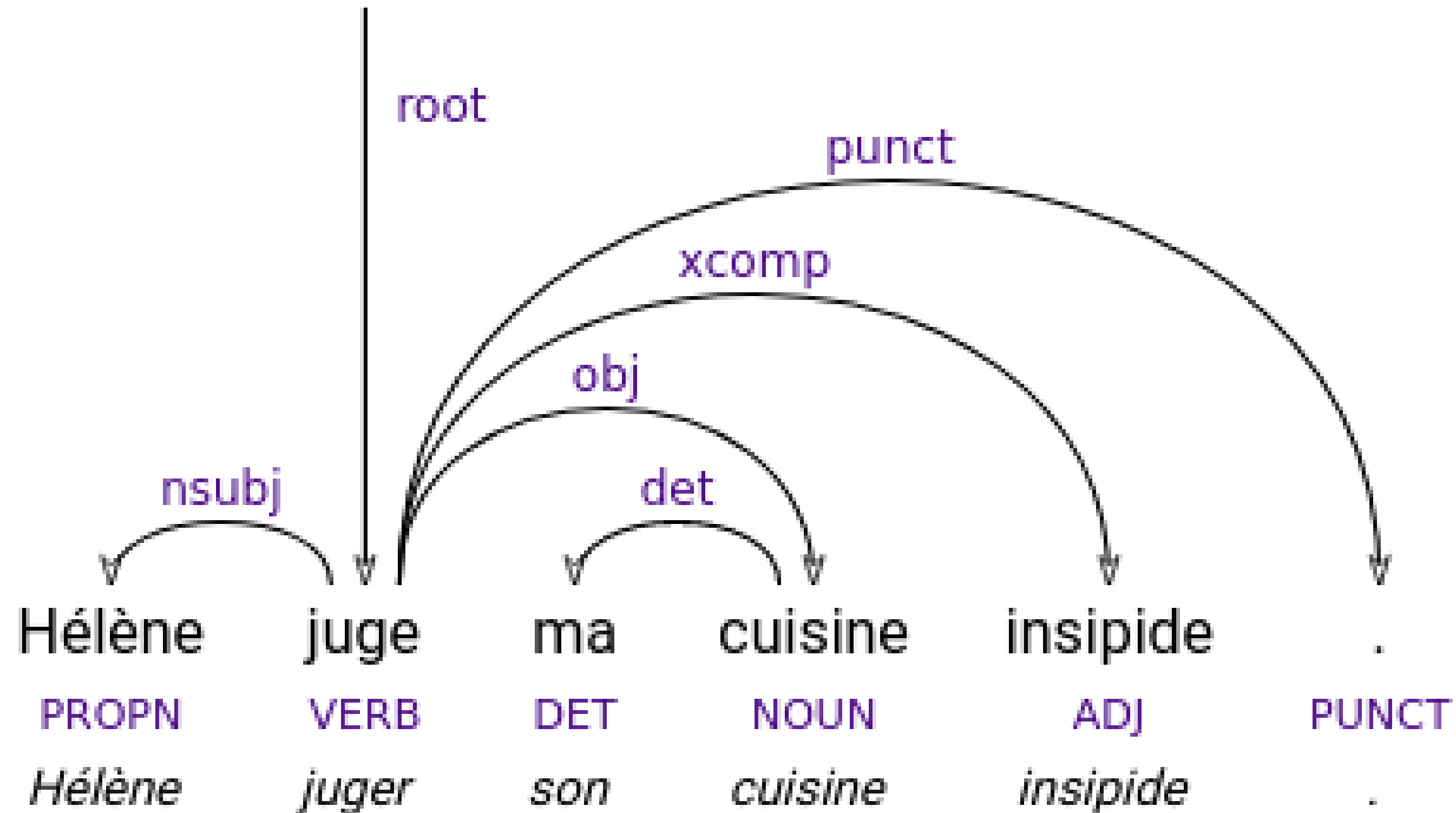
ID: 4    TOKEN: cuisine    LEMMA: cuisine    POS: NOUN    HEAD: 2    DEPREL: obj

ID: 5    TOKEN: insipide    LEMMA: insipide    POS: ADJ    HEAD: 2    DEPREL: xcomp

ID: 6    TOKEN: .    LEMMA: .    POS: PUNCT    HEAD: 2    DEPREL: punct



# Attribut du COD : Stanza



Complément du COD  
et COI:

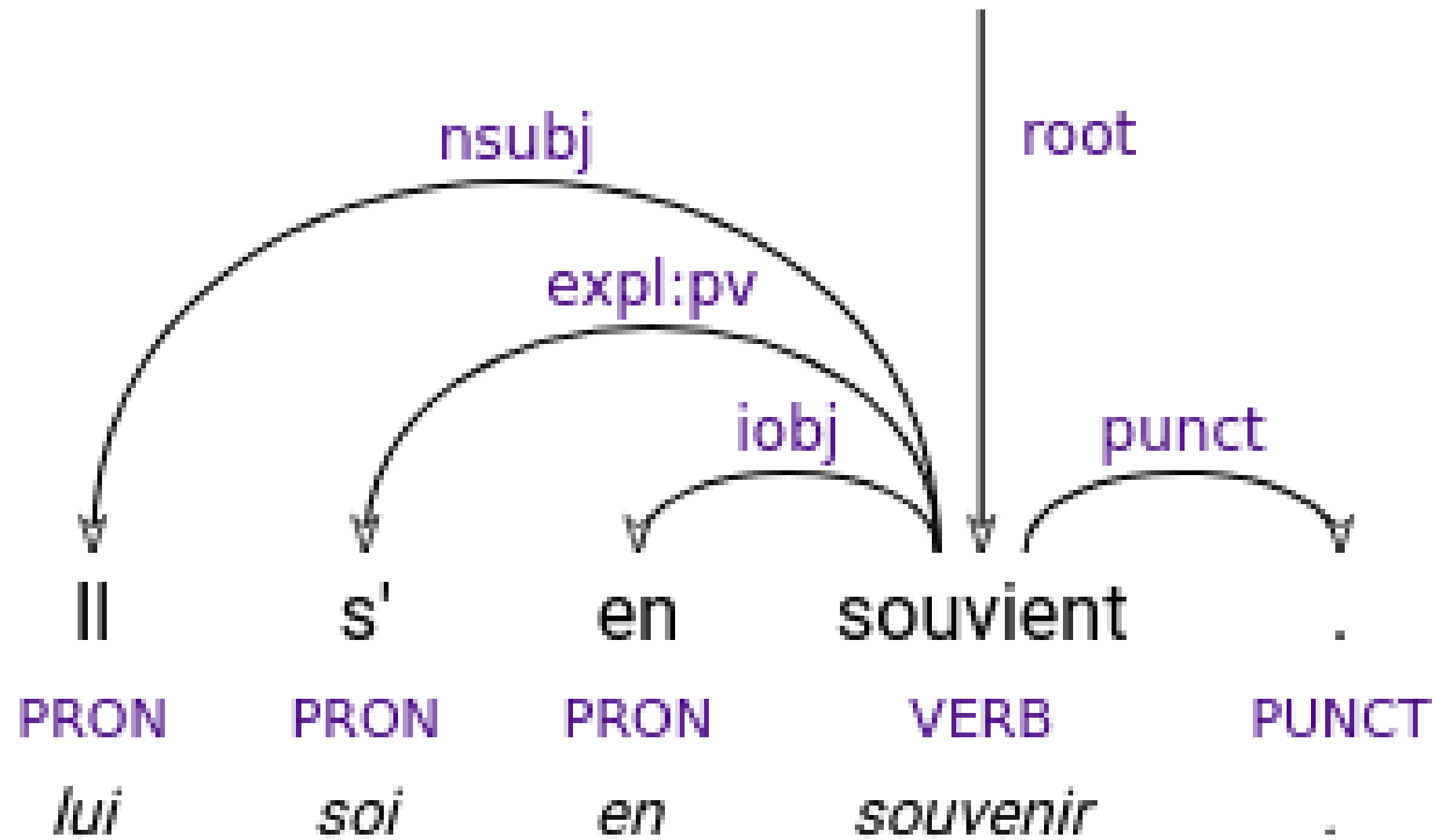
# UDPipe

1	Il	il	PRON	_	Gender=Masc Number=Sing Person=3 PronType=Prs	4	nsubj	_
2	s'	s'	PRON	_	Person=3 PronType=Prs Reflex=Yes	4	expl:pv	_
3	en	en	PRON	✓	Person=3 PronType=Prs	4	✓ iobj	✓ _
4	souvient	souvenir	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	0	root	✓ _
5	.	.	PUNCT	_	_	4	punct	_

Figure 46: Capture d'écran de l'annotation de la phrase "Il s'en souvient." en utilisant la bibliothèque UDPipe.

Complément du COD  
et COI:

# UDpipe



Complément du COD  
et COI:

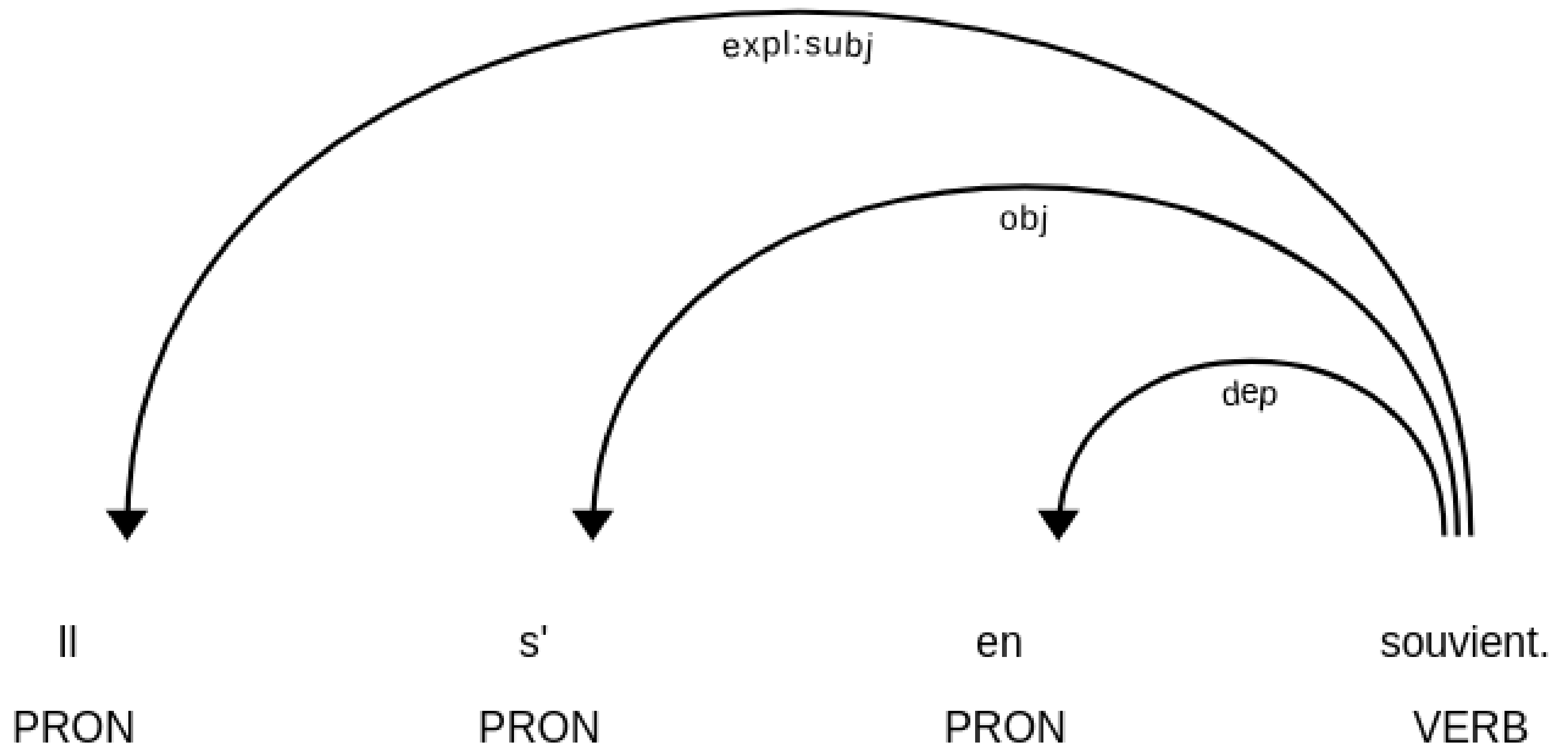
# Spacy

ID: 0	TOKEN: Il	TAG: pronoun	Head: 3	DEPREL: nsubj
ID: 1	TOKEN: s'	TAG: adverb	Head: 3	DEPREL: cop
ID: 2	TOKEN: en	TAG: adposition	Head: 3	DEPREL: case
ID: 3	TOKEN: souvent	TAG: verb	Head: 3	DEPREL: ROOT
ID: 4	TOKEN:	TAG: space	Head: 3	DEPREL: dep

Figure 47: Capture d'écran de l'annotation de la phrase "Il s'en souvient." en utilisant la bibliothèque Spacy.

Complément du COD  
et COI:

# Spacy



Complément du COD  
et COI:

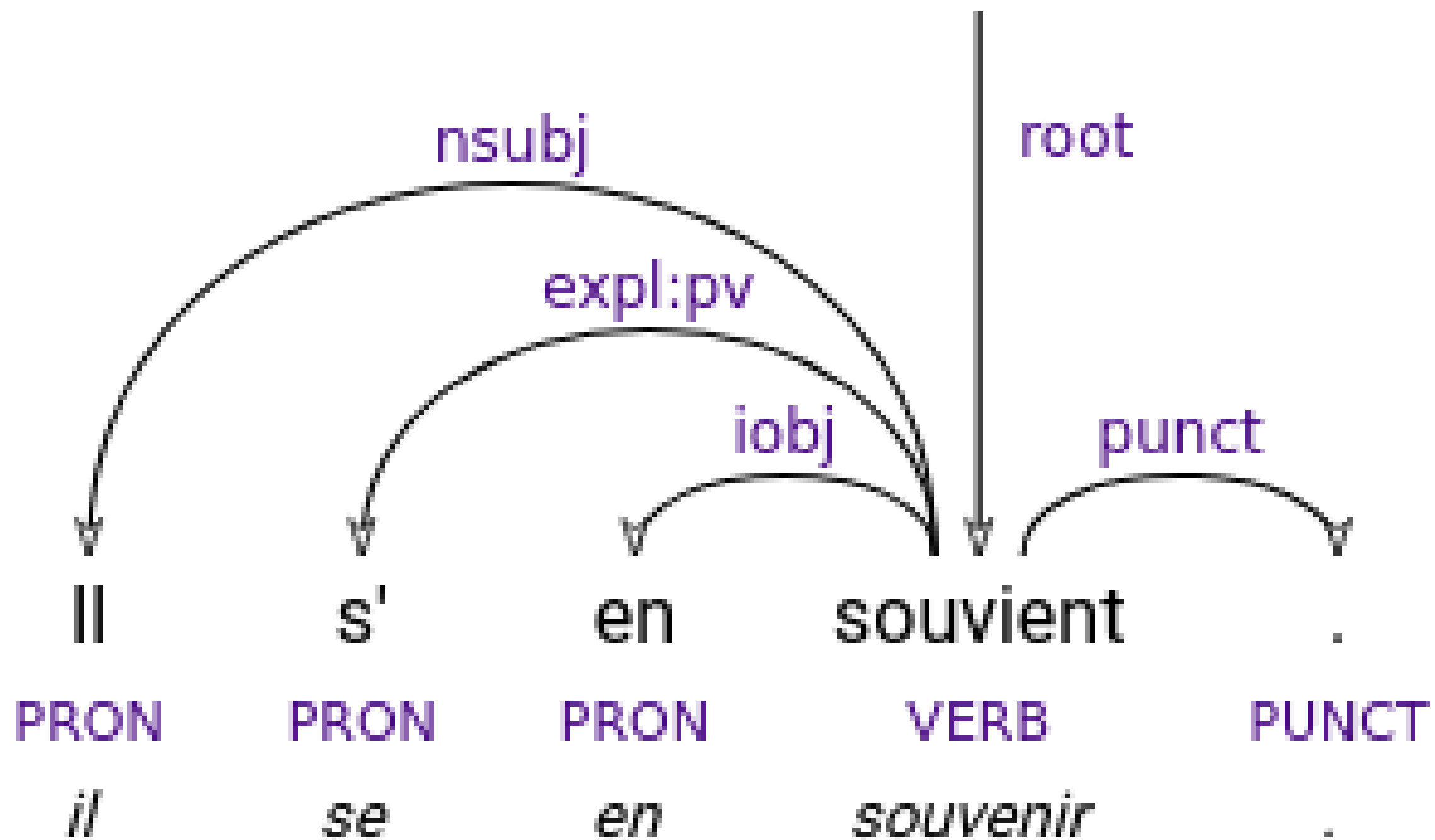
# Stanza

```
Phrase : Il s'en souvient.  
Mot : Il      Lemme : il      Pos : PRON      Head : 4      Relation : nsubj  
Mot : s'      Lemme : se      Pos : PRON      Head : 4      Relation : expl:pv  
Mot : en      Lemme : en      Pos : PRON      Head : 4      Relation : iobj  
Mot : souvient Lemme : souvenir Pos : VERB      Head : 0      Relation : root  
Mot : .      Lemme : .      Pos : PUNCT      Head : 4      Relation : punct
```

Figure 48: Capture d'écran de l'annotation de la phrase "Il s'en souvient." en utilisant la bibliothèque Stanza

Complément du COD  
et COI:

# Stanza



# Comparaison l'évaluation de corpus contenant un attribut du COD

Voici le tableau des résultats sur l'évaluation du corpus portant l'attribut du COD :

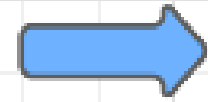
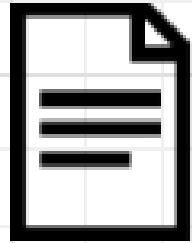
	UDPipe		Spacy		Stanza	
	HEAD	DEPREL	HEAD	DEPREL	HEAD	DEPREL
<i>Hélène juge ma cuisine insipide.</i>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	✓	<b>X</b>
<i>Max a trouvé ce vieux musicien hongrois génial.</i>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
<i>Max a appelé son chien Milo.</i>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
Pourcentage de bons résultats	<b>0 %</b>		<b>0 %</b>		<b>16,6 %</b>	



# Comparaison l'évaluation de corpus contenant un complément du COD/COI

	UDPipe		Spacy		Stanza	
	HEAD	DEPREL	HEAD	DEPREL	HEAD	DEPREL
<i>J'appelle Clément.</i>	✓	✓	✓	✓	✓	✓
<i>J'aime la musique.</i>	✓	✓	✓	✓	✓	✓
<i>J'aime dessiner.</i>	✓	X	✓	X	✓	X
<i>Je me regarde dans le reflet de l'eau.</i>	✓	✓	✓	X	✓	✓
<i>Je te ressemble.</i>	✓	X	X	X	✓	✓
<i>Il s'en souvient.</i>	✓	✓	✓	X	✓	✓
<i>J'y pense.</i>	✓	✓	✓	✓	✓	✓
Pourcentage de bons résultats	85,7 %		64,2 %		92,8 %	

# Réalisation technique



TOKENISATION



EXTRACTION



TRANSFORMATION => JSON

[Retour à la vue d'ensemble](#)

J'appelle  
Clément.

J'

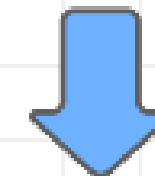
APPELLE

CLÉMENT

.

ID: 1	TOKEN: J'	LEMMA: il	POS: PRON	HEAD: 2	DEPREL: nsubj
ID: 2	TOKEN: appelle	LEMMA: appeler	POS: VERB	HEAD: 0	DEPREL: root
ID: 3	TOKEN: Clément	LEMMA: Clément	POS: PROPN	HEAD: 2	DEPREL: obj
ID: 4	TOKEN: .	LEMMA: .	POS: PUNCT	HEAD: 2	DEPREL: punct

```
question: "Donner la fonction de l'élément figurant en rou  
phrase: "J' appelle **Clément**.",  
reponses: [  
  {  
    texte: "Complément de l'adjectif",  
    correct: false  
  },  
  {  
    texte: "Complément d'agent",  
    correct: false  
  },  
  {  
    texte: "Complément du nom",  
    correct: false  
  },  
  {  
    texte: "COI",  
    correct: false  
  },  
  {  
    texte: "Complément de lieu",  
    correct: false  
  },  
  {  
    texte: "Complément circonstanciel de cause",  
    correct: false  
  },  
  {  
    texte: "COD",  
    correct: true  
  }  
]
```



JAVASCRIPT  
+HTML + CSS

APPLICATION

- **Amélioration du projet**
- **Conclusion**

**Merci pour vos attentions!**

# Questions et réponses.