

FUDIA détecte les interrogatives d'un corpus annoté en dépendances universelles (UD)

Est-ce que l'extraction des interrogatives du français peut-elle être automatisée ?

Valentin D. Richard, *Journées LIFT 2023, Nancy*

1 Introduction

- Interrogatives souvent récupérées à la main – 12 participant·es
- Cohen κ : $\mu = 0,781$

2 Proposition

FUDIA :

- Réécriture de graphe UD
- Par règles

French Interrogative Bank :

- 2.972 phrases
- Étend le FQB

3 Évaluation

- Corpus d'évaluation pré-parsé

- Tâche d'annotation

- Désaccords : enchâssées et marqueurs

4 Résultats

	exa.	pr.	rap.
FUDIA	0,905	0,966	0,770
écrit	0,860	0,917	0,647
oral	0,950	1,00	0,875

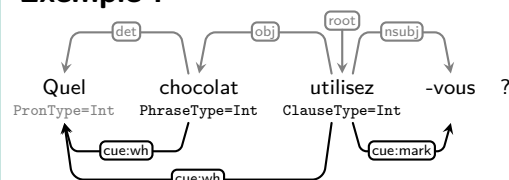
Faux négatifs dus au parseur

- confusion mots interrogatifs vs. mots relatifs ou subordonnants
- ambiguïté du *si*

Corpus UD français :

Écrit : FQB, GSD, ParTUT, PUD, Sequoia
Oral : ParisStories, Rhapsodie

Exemple :



Corpus d'évaluation :

Écrit : ANNODIS, TenNovels, Defrancq
Oral : OFROM, Maya, Defrancq

Erreurs :

- a. mais il fallait répondre à *qui* (DO)
- b. Nous ne voulons pas savoir *comment* les faits se sont déroulés [...] (DE)
- c. [...] l'AIEA n'était pas en mesure de vérifier s'il y a eu détournement ou non de matériel nucléaire [...] (DE)
- d. ce qui me préoccupe c'est que *va-t-on* faire avec XXX (DO)

Statistiques FIB

