



HACKATAL 2023

DÉTECTION DE CITATIONS

Aurélie Nomblot et Yagmur Ozturk



5 et 6 juin 2023

Méthodologie : utilisation de règles (RegExr)



- maîtrisé
- rapide
- efficace
- performant sur un corpus modéré
- fun

Le type de la proposition contenant les paroles rapportées

Le verbe le plus fréquemment utilisé pour introduire les paroles des personnages est le verbe *dire* ; mais c'est un verbe « neutre » : il donne peu de renseignements. De très nombreux verbes peuvent le remplacer.

En voici quelques-uns, classés selon leur sens et leur emploi :

- Certains verbes annoncent :

- **une phrase déclarative** : *dire, affirmer, déclarer, répéter, annoncer, ajouter, poursuivre, reprendre, préciser, confirmer, apprendre, raconter, noter, conclure, révéler, dévoiler, expliquer...*
- **une phrase interrogative** : *demande, questionner, interroger, s'enquérir, s'informer, se renseigner...*
- **une phrase exclamative** : *s'écrier, s'exclamer, se récrier, s'indigner...*
- **une deuxième prise de parole ou une réponse** : *répondre, répliquer, rétorquer, ajouter, répéter, riposter, repartir, récriminer...*
- **une phrase injonctive** : *ordonner, exiger, commander, recommander, conseiller...*

- Quelques verbes indiquent que le personnage se parle à lui-même, qu'il pense : *se dire, songer, penser, réfléchir, raisonner, spéculer...*

- D'autres renseignent sur l'humeur du personnage qui parle, sur la manière dont il s'exprime : *marmonner, bougonner, grommeler, pleurnicher, bredouiller, bégayer...*

	Discours direct	Discours indirect
Indices de personne	Je, tu Nous, vous Mon, ton, nos, vos	Il, elle Ils, elles Son, sa, ses, leurs
Adverbes de lieu	Ici Là-bas	Là À cet endroit là
Adverbe de temps	Hier Aujourd'hui Demain	La veille Ce jour-là Le lendemain
Temps des verbes	Présent Futur Imparfait Passé Composé Futur Antérieur	Imparfait Conditionnel présent Imparfait ou Plus-que-parfait Plus-que-parfait Conditionnel passé

Typologie du discours	Marqueurs	Exemples du corpus	Expressions Régulières	Performance estimée
Discours direct	..., « ... » + dit-il / ajouta-t-elle...	Avec le 49.3, « nous avons atteint notre objectif, ce texte n'a aucune légitimité, on a raison de se révolter », a-t-il ajouté.	[^"«»'«„“(.*?)[^"«»'«„“(,?) (a avait) (dit pensé ...) (\b[A-Z][a-zA-zââçéèëëïïôœùû'-]+\b)? (\b[A-Z][a-zA-zââçéèëëïïôœùû'-]+\b)?.	99,9%
	... dit : « ... ».	l'Insoumis a appelé au calme : « Notre force principale c'est pas quatre poubelles qui brûlent, c'est la force du nombre ».	[A-Za-zÀ-ÿ\s]+(a avait) (dit pensé affirmé déclaré répété...) ??:[^.]+	99,9%
	Selon lui / d'après lui, « ... ».	Selon lui, « c'est le prix à payer pour avoir une alimentation qui soit saine et durable ».	(Selon D'après) (lui elle),? [^"«»'«„“(.*?)[^"«»'«„“(99,9%
Discours indirect	Elle dit que...	elle a répondu: Oui, il y a une mission au niveau des ministres des Affaires étrangères, il y en plusieurs en ce moment qui se rendent en Tunisie .	[A-Za-zÀ-ÿ\s]+(a avait) (dit pensé affirmé déclaré répété ajouté raconté noté conclu révélé dévoilé...) ??:[^.]+	?
	..., assure-t-elle.	On n'ira pas comme ça , assure-t-elle	[^.] +(a affirmé assure-t-elle fulmine a-t-il déclaré)((\b[A-Z][a-zA-zÿ'-]+\b)? (\b[A-Z][a-zA-zÿ'-]+\b)?)?.	?
	... reconnaît ...	Il faudra une négociation , reconnaît...	[A-Za-zÀ-ÿ\s]+,? (,(?(reconnaît ...riposte ordonne commande recommande conseille prononce))([A-Za-z]+]?)	?

Expressions Régulières

[""](.*)[""](.?)
(a|avait) (dit|pensé|...)
(\b[A-Z][a-zA-Zâççêê
ëïïôœùù'-]+ \b)?
(\b[A-Z][a-zA-Zâççêêê
ëïïôœùù'-]+ \b)?.

[A-Za-zÀ-ÿ\s]+(a|avait)
(dit|pensé|affirmé|décl
aré|répété...) ??[^\.]+

(Selon|D'après)
(lui|elle)? [""]
« . »(.*)[""]

[A-Za-zÀ-ÿ\s]+(a|avait)
(dit|pensé|affirmé|décl
aré|répété|ajouté||raco
nté|noté|conclu|révélé|
dévoilé...) ??[^\.]+

[^\.]+(a
affirmé|assure-t-elle|ful
mine| a-t-il
déclaré)((\b[A-Z][a-zA-
ÿ'-]+ \b)? (\b[A-Z][a-zA-
ÿ'-]+ \b)?).

[A-Za-zÀ-ÿ\s]+,?
(,?(reconnaît|...riposte|
ordonne|commande|re
commande|conseille|p
rononce))([A-Za-z]+
)?

opt



-]+(a (dit|proposé|rélevé|répondu|per

163 matches (9.2ms)

grédients sont en place pour que la
outes les actualités du site
ansmettons cette actualité
gie programmatique sur le site et

s ne produisent pas des travaux de
Emmanuel Macron le Président
prononcé à l'Elysée en prélude
Gabon, en Angola, au Congo et en RD
l Macron a demandé aux entreprises
une profonde humilité face à ce qui
in au profit de nuances comme la

- get_content(corpus)

Extraire le contenu

Observer des formulations

Tester les expressions régulières

- extract_quote(liste_de_contenu)

Ajout des expressions régulières

Recherche de match (span)

Sortie : dictionnaire {index : [[start, end, "Citation"], ...], ...}

- add_json(entrée, sortie, dictionnaire)

lecture du json d'entrée

insertion du span

écriture du json de sortie

```
"spans": {  
  "0": [  
    [  
      564,  
      583,  
      "Citation"  
    ],  
    [  
      657,  
      676,  
      "Citation"  
    ],  
    [  
      900,  
      909,  
      "Citation"  
    ],  
    [  
      1818,  
      1829,  
      "Citation"  
    ]  
  ]  
},
```

- get_content(corpus)

Extraire le contenu

Observer des formulations

Tester les expressions régulières

- extract_quote(liste_de_contenu)

Ajout des expressions régulières

Recherche de match (span)

Sortie : dictionnaire {index : [[start, end, "Citation"], ...], ...}

- add_json(entrée, sortie, dictionnaire)

lecture du json d'entrée

insertion du span

écriture du json de sortie

```
"spans": {  
  "0": [  
    [  
      564,  
      583,  
      "Citation"  
    ],  
    [  
      657,  
      676,  
      "Citation"  
    ],  
    [  
      900,  
      909,  
      "Citation"  
    ],  
    [  
      1818,  
      1829,  
      "Citation"  
    ]  
  ]  
},
```


- get_content(corpus)

Extraire le contenu

Observer des formulations

Tester les expressions régulières

- extract_quote(liste_de_contenu)

Ajout des expressions régulières

Recherche de match (span)

Sortie : dictionnaire {index : [[start, end, "Citation"], ...], ...}

- add_json(entrée, sortie, dictionnaire)

lecture du json d'entrée

insertion du span par index de contenu


écriture du json de sortie

```
"spans": {  
  "0": [  
    [  
      564,  
      583,  
      "Citation"  
    ],  
    [  
      657,  
      676,  
      "Citation"  
    ],  
    [  
      900,  
      909,  
      "Citation"  
    ],  
    [  
      1818,  
      1829,  
      "Citation"  
    ]  
  ]  
},
```

HackaTAL 2023 - Détection de Citations.

Leader board.

2023-06-06 09:45:58.411686



	name	MAX(f)
0	HackaTeam	1.0000
1	FlowBERT	0.6569
2	Team 6	0.4301
3	Alèd	0.2090

Hacka

Leader l

2023-06-06 09:

0

1

2

3



VICTORY!

MAX(f)

1.0000

0.6569

0.4301

0.2090

Perspectives d'améliorations

- Ajouter du POS tagging et de REN
 1. [A-Za-zÀ-ÿ\s]+(a|avait) (dit|pensé|affirmé|déclaré|répété...) ?:[^.]+
 2. Emmanuel Macron **EN** +(a|avait) (dit|pensé|affirmé|déclaré|répété...) ?:[^.]+
- Analyse approfondie sur la position et la forme des propositions (proposition introductive et proposition rapportée)

“Wen the reporting clause located within or after the reported clause, subject and verb may revers positions [...]”¹



¹ Ralf Krestel, Sabine Bergler, and René Witte. 2008. [Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).



Aurélie Nomblot : aurelie.nomblot@gmail.com

Yagmur Ozturk : yagmur.ozturk@edu.univ-fcomte.fr