

Les chaines anaphoriques avec un mot interrogatif ne sont pas toutes (bien) annotées dans ANCOR

Évaluation d'un corpus annoté en anaphores : le cas des chaines contenant un mot interrogatif

Valentin D. Richard, *Journées LIFT2 2025, Paris*

1. Chaine de mentions : antécédent, coréférence, anaphore associative

(1) Les enfants_i sont dehors.
Ils_i jouent au ballon.
Paul_{i'} est le goal.

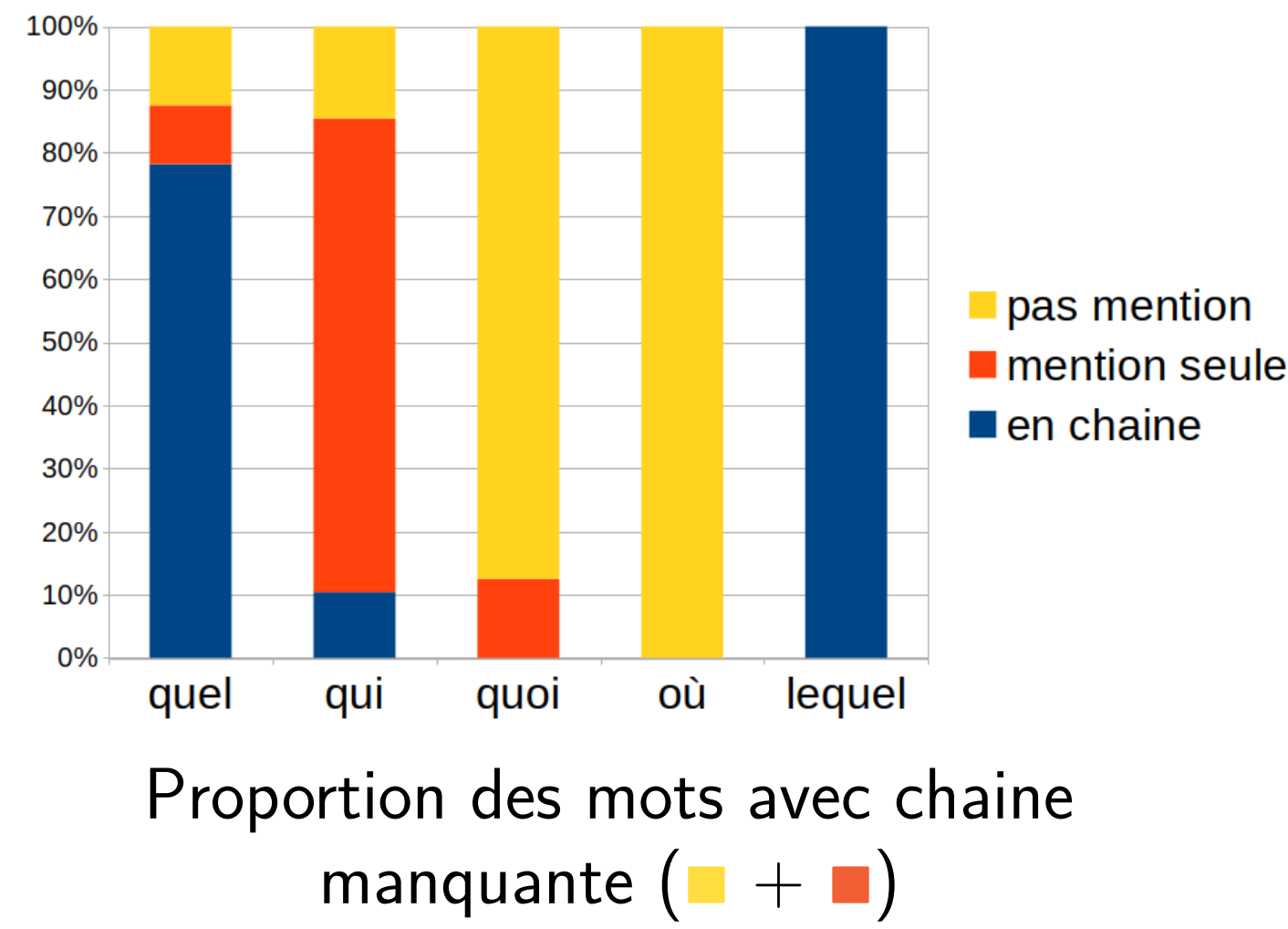
ANCOR : corpus oral annoté en anaphores

2. Objectif : Évaluer quantitativement et qualitativement les chaines anaphoriques qui contiennent un mot interrogatif

3. Méthode :

- (i) Parsage d'ANCOR en UD (ArboratorGrew + FUDIA) et extraction des mots interrogatifs
- (ii) Annotation manuelle des chaines contenant un mot interrogatif

4. Résultats : ci-contre



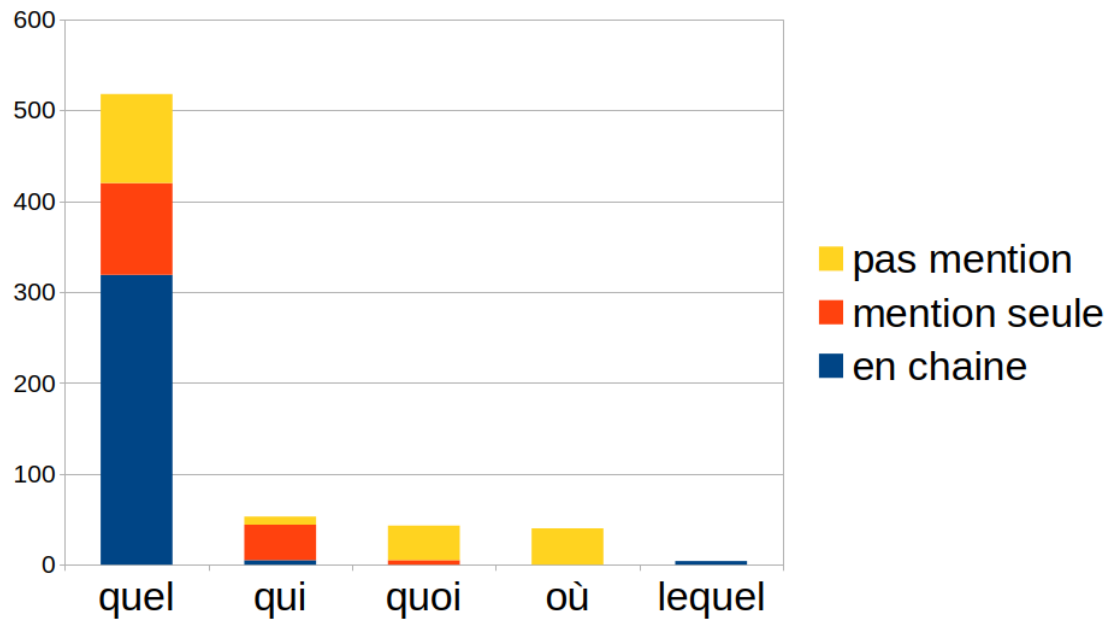
5. Analyse :

- chaines avec *quel* bien annotées
- beaucoup de mentions et chaines avec *qui*, *quoi* et *où* manquantes et annotations peu cohérentes

6. Conclusion : Besoin d'un guide d'annotation des anaphores unifié pour les mots interrogatifs

Muzerelle, J., Lefevre, A., Schang, E., Antoine, J.-Y., Pelletier, A., Maurel, D., Eshkol, I., & Villaneau, J. (2014). ANCOR_Centre, a large free spoken French coreference corpus : Description of the resource and reliability measures. LREC'14 (p. 843-847). ELRA

Mots interrogatifs trouvés :



Annoté dans ANCOR :

Question en *quel* \bar{N}_j et réponse en c_j 'est $X_{j'}$
(2) SPK2: et vous personnellement dans quelle matière_j vous aimeriez que vos enfants soient forts à l'école ?
SPK1: c_j 'est les mathématiques_{j'} 096_C-1

Manquants :

- Pas en mention, mais spécifique :
(3) SPK1: la médiathèque c'est quoi_k c'est pour les disques
SPK2: pour les disques
SPK1: alors c_k 'est la discothèque_{k'} 1AG0618
- Pas en mention, réponse non exhaustive :
(4) SPK3: à Orléans y a aucune distraction
SPK1: où_k est-ce que vous allez alors ? 542_C-1
SPK3: on va même à Beaugency_{k'} y a y a déjà des dancing Beaugency_{k'} Neuville aux bois_{k''}

Incohérences : domaine

- (5) et en général dans votre famille_j qui_j est-ce qui se sert le plus souvent du dictionnaire ou des dictionnaires ? C02_ESL0_003_C

Difficultés : réponses conditionnelles

- (6) SPK2: et qui_k est-ce qui remplit les papiers administratifs chez vous ? [...]
SPK1: euh la feuille d'impôts c_k 'est évidemment moi_{k'} [...] mais par contre les feuilles de sécurité sociale c_k 'est ma femme_{k''} 012_C-3

Code source :

