



HAL
open science

Outiller l'étude des chaînes de référence dans des écrits scolaires

Martina Barletta, Claude Ponton

► **To cite this version:**

Martina Barletta, Claude Ponton. Outiller l'étude des chaînes de référence dans des écrits scolaires. Journées de Linguistique des Corpus, Jul 2023, Grenoble, France. hal-04153289

HAL Id: hal-04153289

<https://hal.science/hal-04153289>

Submitted on 6 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Outiller l'étude des chaînes de référence dans des écrits scolaires



Martina Barletta^{1,2}, Claude Ponton¹

¹Laboratoire LIDILEM, Université Grenoble Alpes (France),

²Doctorat en Education dans la Société Contemporaine, Université Milan-Bicocca (Italie)

Bibliographie



Objectif

- Description linguistique de la cohérence et de la cohésion textuelles, et son évolution (en particulier, chaînes de référence) dans *Scolinter*, corpus d'écrits scolaires du primaire en français, italien et espagnol



Problématique

- En quoi le TAL peut-il assister cette description linguistique ?
 - Proposer un modèle d'annotation adapté aux chaînes de référence dans les écrits scolaires
 - Développer une méthodologie et un outillage pour la pré-annotation de ces chaînes dans *Scolinter*

Corpus français et coréférence

Corpus	Année	Taille	Genres	Phénomènes annotés
ANCOR	2013	487 000 mots	parole spontanée transcrite	chaînes de coréférence ou anaphore
<i>democrat</i>	2019	560 000 mots, 58 textes	écrits narratifs et autres genres variés	chaînes de référence
<i>Corpus RésolCo</i>	2021	97 798 mots, 407 textes	écrits scolaires	continuité référentielle sur les entités provoquées par la consigne

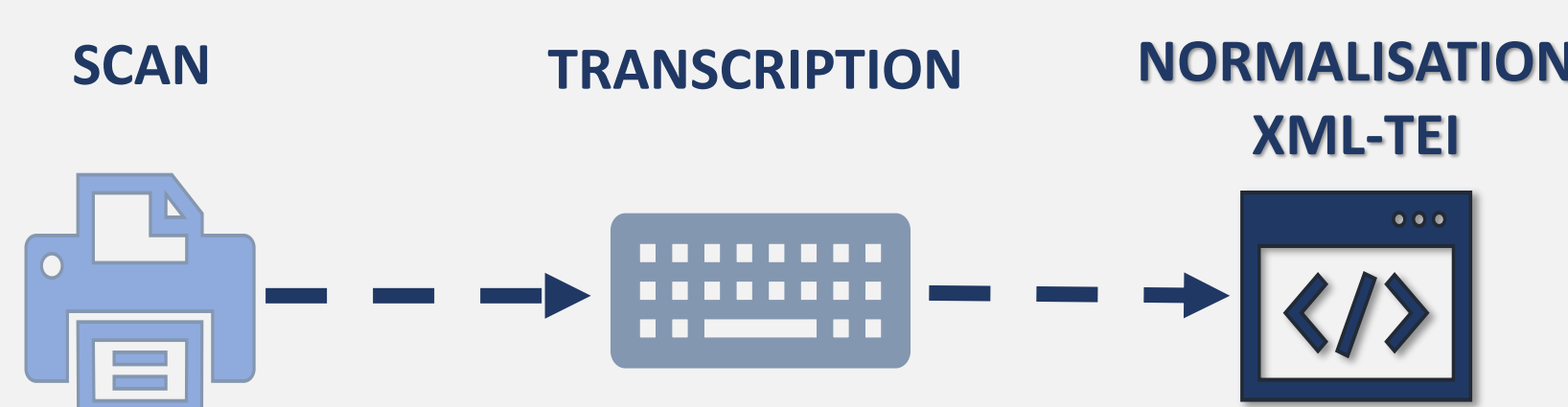
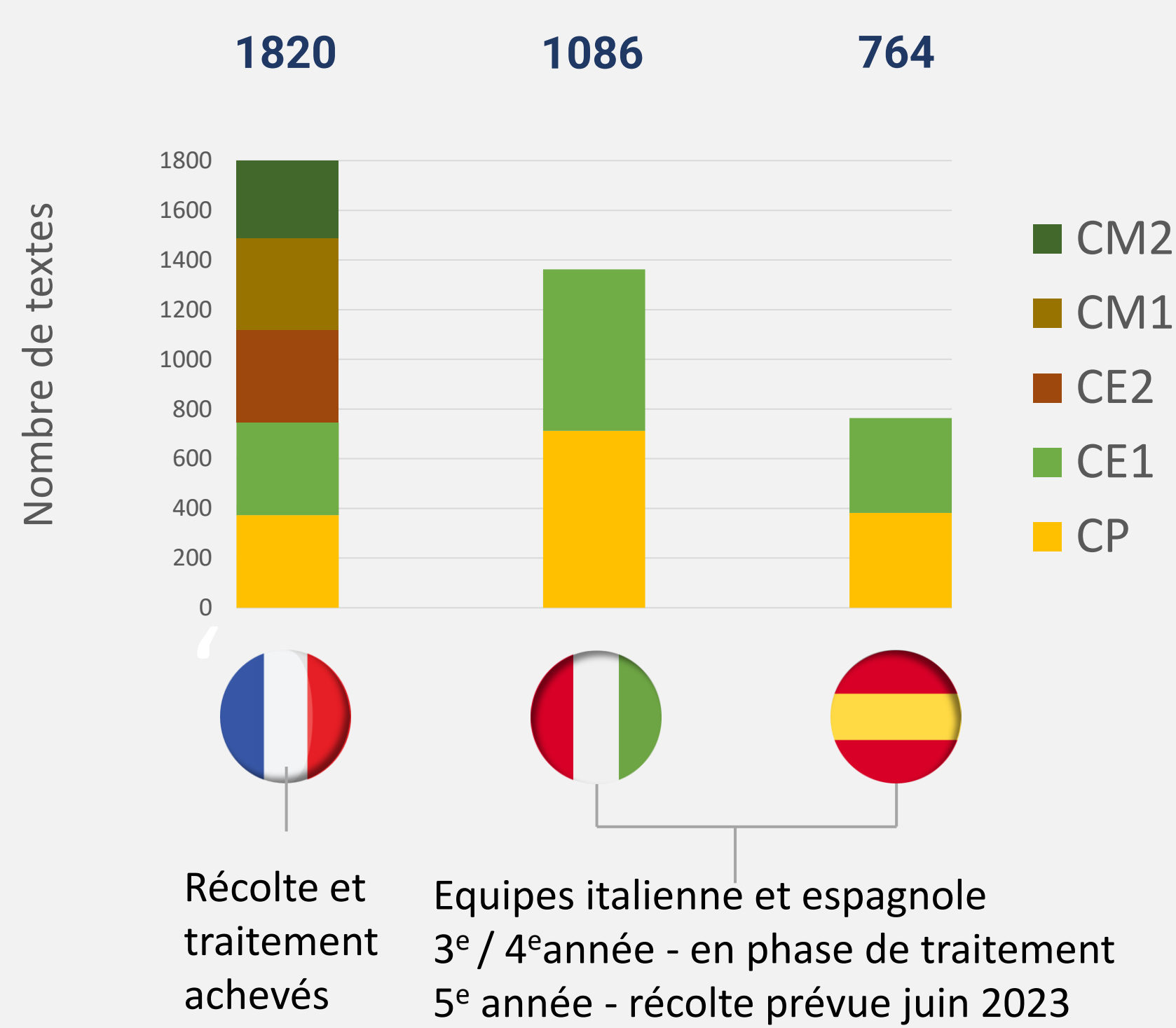
Chaîne de référence

ensemble d'au moins trois expressions faisant référence à la même entité, appartenant à l'univers du texte (Schneidecker, 1997)

Corpus *Scolinter*



Production d'un récit à partir d'images, du CE1 au CM2

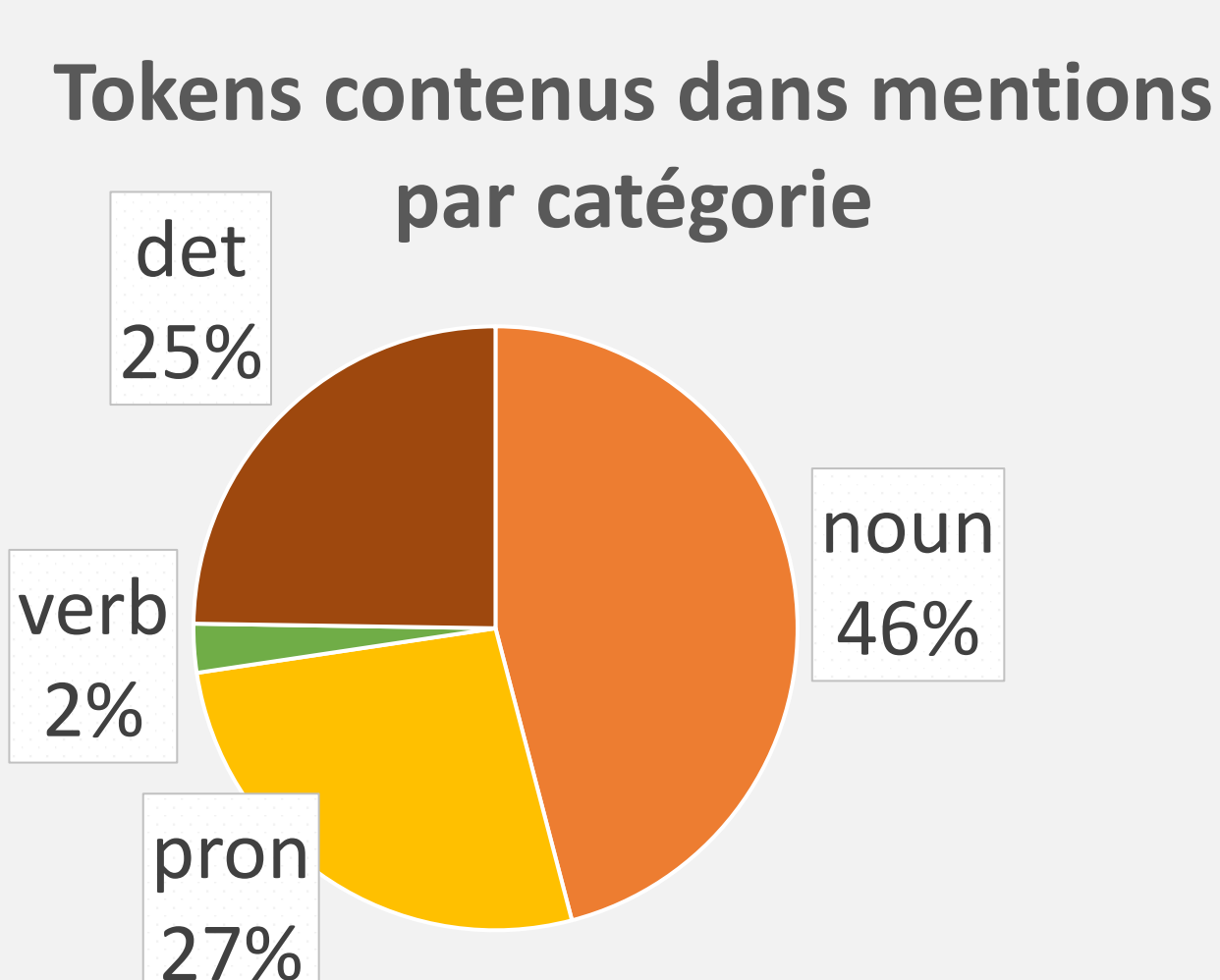
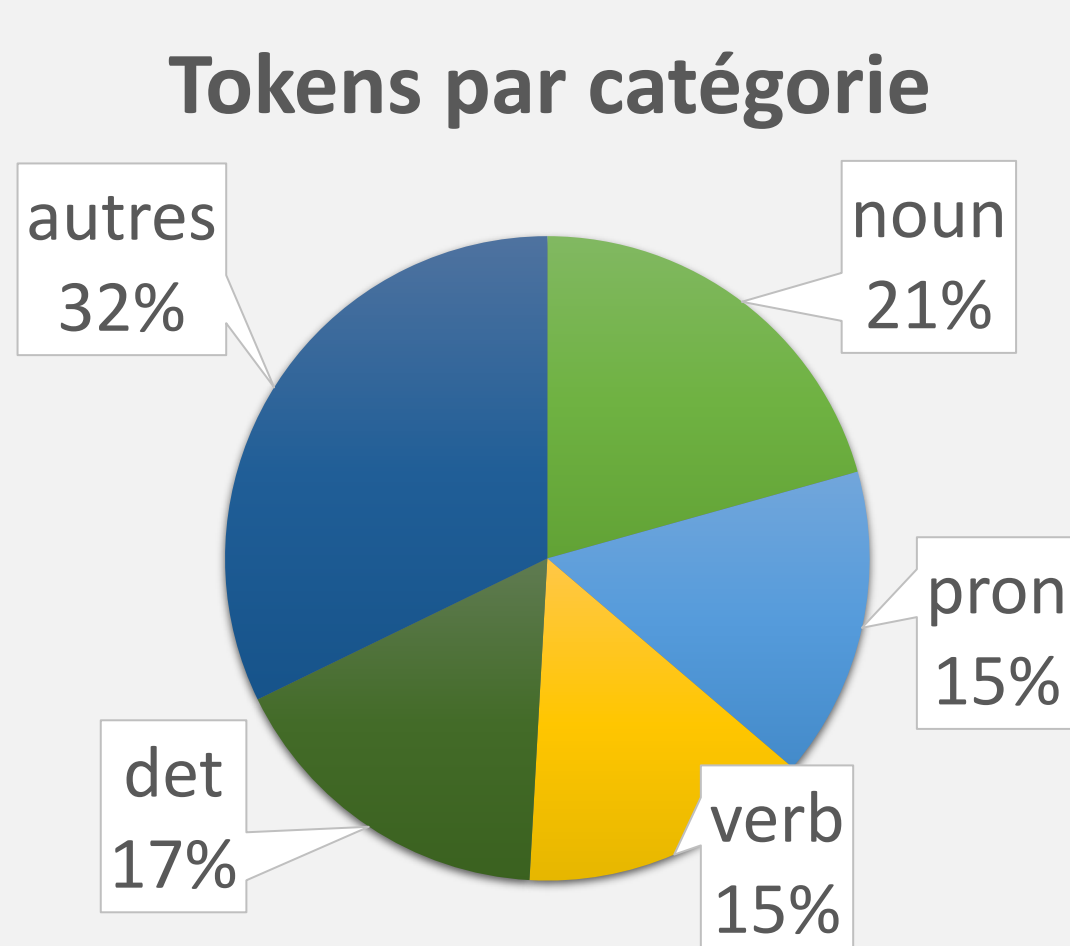
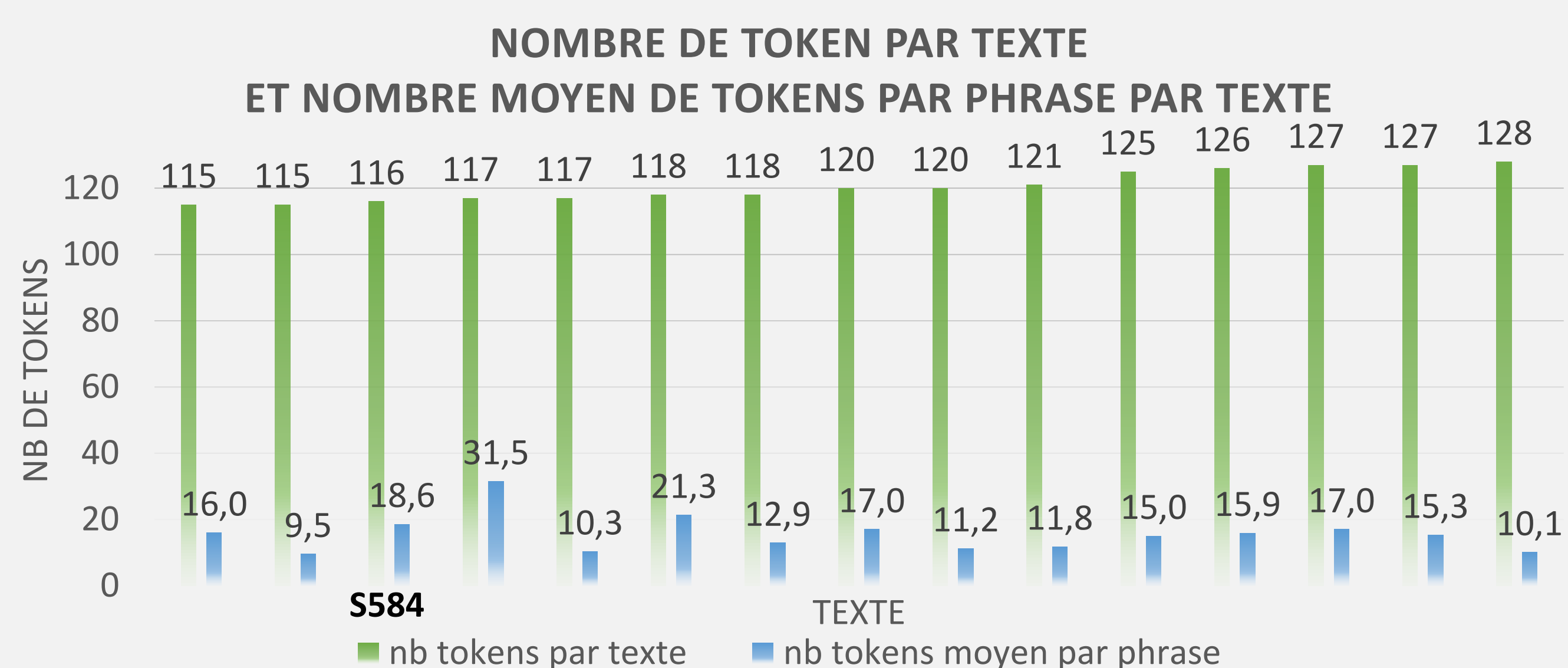


Avancée du travail

- Modèle d'annotation inspiré de *Democrat* et de *RésolCo* – en cours de finalisation
- Types de mentions visées
 - Nominales
 - Pronominales
 - Déterminants possessifs
 - Verbes (en absence du sujet)
- Outil déjà existant pour la pré-annotation des chaînes de coréférence – *DeCorScol* (Barletta, 2022)
- Création d'un corpus de référence pour l'évaluation des outils (en cours)

Caractérisation du corpus de travail

15 textes CE2, annotés selon le modèle RésolCo/Democrat



Exemple d'annotation

Le chat et la sorcière

C'est l'histoire d'un chat très malin qui vit dans un château, tous les soirs le chat sort pour aller au village. un soir, il se perd dans la forêt croit qu'il allait au village, il prend peur, court dans tous les sens. À un moment il trouve une cabane il entrait, se coucha sur le canapé.

Le lendemain il tombe nez à nez avec la sorcière du village, la sorcière prend le chat et le transforma en chat noir et elle dit « joue avec moi » le chat étonné approche d'elle et la caresse. Maintenant chaque soir ils se retrouvent dans la forêt pour jouer. Fin.

S584 – CE2