



# Annotation de la co-référence

Journée d'étude organisée par

L.-M. Ho-Dac, J. Rebeyrolle, C. Garcia-Debanc et M.-P. Jacques



# Annotation de la co-référence



Corpus ANCOR

dem  
ocrat

ANR Democrat

**langue française**  
195 septembre 2017  
Les chaînes de référence en corpus

Catherine Schneedecker,  
Julie Gilkman & Frédéric Landragin

Emmanuel Baumer  
Delphine Bernhard  
Bernard Combettes  
Thomas François  
Nuria Galà  
Céline Guiller-Barbance  
Rafika Khodja  
Anne-Laure Ligistat  
Vanessa Oby  
Bénédicte Pincemin  
Amalia Todorcu

ARMAND COLIN

programme de la journée en ligne



Journée d'étude

## Annotation de la coréférence

organisée par Mai Ho-Dac, Josette Rebeyrolle, Claudine Garcia-Debanc et Marie-Paule Jacques  
CLLE (UMR 5263) – Université de Toulouse Jean-Jaurès

**Jeu**di 19 novembre 2020  
**Maison de la Recherche - Salle D29**

Cette journée propose une confrontation d'approches variées de l'annotation de la coréférence dans des corpus de textes diversifiés. Elle sera l'occasion de discuter notamment des questions suivantes :

- Choix du modèle d'annotation pour identifier et délimiter les expressions référentielles et pour relier celles qui renvoient à un même référent dans un discours, c'est-à-dire aux « chaînes de coréférents »
- Mise en place des schémas et procédures d'annotation : pré-annotations automatiques, lourdeur de la tâche de saisie et/ou de correction manuelle, possibilité de s'aider d'outils de TAL
- Annotation de la coréférence dans des textes dont les scripteurs ne sont pas des experts mais des apprenants, comme dans les écrits scolaires.

Conférence invitée : **Étude des chaînes de référence en français : le projet ANR Democrat**

Frédéric Landragin (*Laboratoire LATTICE, ENS, UMR 8094*)

Résumé :

Une chaîne de référence – ou ensemble des expressions référentielles qui portent sur un même référent – est un objet linguistique dont l'étude implique de nombreux aspects : distinction entre l'accès à un référent et l'évocation (non référentielle) d'un référent ; délimitation des expressions référentielles ; caractérisation des successions des expressions référentielles, par exemple des transitions d'un référent à un autre ; mise au jour de typologies aussi bien pour les expressions référentielles que pour les référents et les chaînes elles-mêmes. Ces aspects ont fait l'objet de discussions dans le cadre du projet ANR Democrat (2016-2020). Nous présentons les grandes lignes des modélisations linguistiques et les grands principes des analyses quantitatives qui y ont été effectuées, ainsi que le corpus qui a été constitué et annoté par les membres du projet. Ce corpus a servi de base de travail non seulement pour les modélisations et les analyses, mais aussi pour le développement de fonctionnalités d'annotation et d'interrogation de corpus, ainsi que pour le développement de systèmes de TAL visant à détecter automatiquement des chaînes de référence dans le texte tout-venant.



## Programme de la journée

9h30-9h40	Accueil et ouverture de la journée
9h40-10h10	<b>De Annodis à É:Calm : d'un guide d'annotation de la coréférence à l'autre</b> Lydia-Mai HO-DAC, Silvia FEDERZONI, Josette REBEYROLLE, Claudine GARCIA-DEBANC <i>Laboratoire CLLE, équipe ERSS, UMR 5263, Université de Toulouse Jean Jaurès</i>
10h10-10h40	<b>Annoter des chaînes de référence en corpus d'apprenants pour évaluer des compétences textuelles</b> Luca PALLANTI & Marie-Paule JACQUES <i>Laboratoire LIDILEM, Université Grenoble Alpes</i>
10h40-11h10	PAUSE
11h10-11h40	<b>Chaînes et maillons dans ANNODIS : référents humains versus référents non-humains</b> Silvia FEDERZONI <i>Laboratoire CLLE, équipe ERSS, UMR 5263, Université de Toulouse Jean Jaurès</i>
11h40-12h10	<b>Annoter des textes d'élèves de 9 à 16 ans pour élaborer une cartographie des compétences en cohésion textuelle</b> Claudine GARCIA-DEBANC <i>Laboratoire CLLE, équipe ERSS, UMR 5263, INSPÉ Toulouse Occitanie-Pyrénées</i>
12h10-14h00	Pause déjeuner
14h00-15h00	<b>Étude des chaînes de référence en français : le projet ANR Democrat</b> Frédéric LANDRAGIN <i>Laboratoire LATTICE, ENS, UMR 8094</i>







# Plan

- 1 Projets ANR AnnoDis et E :Calm
  - AnnoDis vs. E :Calm
- 2 Collecte des données : la consigne ResolCo
- 3 Annotation des continuités référentielles dans les données RésolCo
  - Des données primaires aux données annotables
  - Une annotation sélective et adaptée
- 4 Résultats attendus et à venir

# Plan

- 1 Projets ANR AnnoDis et E :Calm
  - AnnoDis vs. E :Calm
- 2 Collecte des données : la consigne ResolCo
- 3 Annotation des continuités référentielles dans les données RésolCo
- 4 Résultats attendus et à venir



[Péry-Woodley et al., 2011]

Confronter différentes approches et modèles de l'organisation discursive

- Constituer une ressource annotée permettant des analyses qualitatives et quantitatives
- Mise à disposition d'outils pour la recherche
- Annoter un corpus diversifié de français écrit (standard) en structures discursives :
  - Unités élémentaires et relations rhétoriques `_rr`
  - Structures multi-échelles `_me` : structures énumératives et chaînes topicales



[Doquet et al., 2017]

Dresser une cartographie des compétences d'écriture au fil de la scolarité

- Constituer une ressource annotée permettant des analyses qualitatives et quantitatives
- Mise à disposition d'outils pour la recherche et l'enseignement
- Annoter des productions d'élèves (écrit non standard) au niveau de l'orthographe et de la cohérence :
  - Unités élémentaires et relations rhétoriques
  - Chaînes de référence



[Péry-Woodley et al., 2011]

Confronter différentes approches et modèles de l'organisation discursive

- Constituer une ressource annotée permettant des analyses qualitatives et quantitatives
- Mise à disposition d'outils pour la recherche
- Annoter un corpus diversifié de français écrit (standard) en structures discursives :
  - Unités élémentaires et relations rhétoriques **\_rr**
  - Structures multi-échelles **\_me** : structures énumératives et **chaînes topicales**



[Dquet et al., 2017]

Dresser une cartographie des compétences d'écriture au fil de la scolarité

- Constituer une ressource annotée permettant des analyses qualitatives et quantitatives
- Mise à disposition d'outils pour la recherche et l'enseignement
- Annoter des productions d'élèves (écrit non standard) au niveau de l'orthographe et de la cohérence :
  - Unités élémentaires et relations rhétoriques
  - **Chaînes de référence**

# Processus +/– commun de constitution



**me** [Péry-Woodley et al., 2017]

- 1 Collecte de 3 sous-corpus de textes expositifs publiés
 

WIKI	articles encyclo.	Wikipedia
GEOP	rapports de géopo.	IFRI
LING	articles académiques	CMLF
- 2 Encodage de la structure de document (format XML TEI-P5)
- 3 Étiquetage morphosyntaxique et analyse syntaxique automatique
- 4 Prémarquage d'indices pour faciliter (aiguiller ?) l'annotation
- 5 Annotation descendante "naïve", sonder l'intuition des locuteurs



- 1 Collecte de copies produites en classes auprès d'élèves de primaire, collège, lycée et université
- 2 Digitalisation des textes (format XML TEI-P5)
- 3 Normalisation orthographique
- 4 Étiquetage morphosyntaxique et analyse syntaxique automatique
- 5 Prémarquage d'indices pour faciliter (aiguiller ?) l'annotation
- 6 Annotation ascendante "experte", appliquer un modèle linguistique



# Processus +/– commun de constitution



**me** [Péry-Woodley et al., 2017]

- 1 Collecte de 3 sous-corpus de textes expositifs publiés
 

WIKI	articles encyclo.	Wikipedia
GEOP	rapports de géopo.	IFRI
LING	articles académiques	CMLF
- 2 Encodage de la structure de document (format XML TEI-P5)
- 3 Étiquetage morphosyntaxique et analyse syntaxique automatique
- 4 Prémarquage d'indices pour faciliter (aiguiller ?) l'annotation
- 5 **Annotation descendante** "naïve", sonder l'intuition des locuteurs



- 1 Collecte de copies produites en classes auprès d'élèves de primaire, collège, lycée et université
- 2 Digitalisation des textes (format XML TEI-P5)
- 3 Normalisation orthographique
- 4 Étiquetage morphosyntaxique et analyse syntaxique automatique
- 5 Prémarquage d'indices pour faciliter (aiguiller ?) l'annotation
- 6 **Annotation ascendante** "experte", appliquer un modèle linguistique

# Annotation "descendante" ⇓ vs. ⇑ "ascendante"



me

- ⇓ des structures
  - chaînes topicales
  - CR de "haut niveau", segments de texte (> 2 phrases)
  - organisés autour d'un référent topical commun
- ⇓ aux indices
  - indices de surface signalant la structure ("maillons")



- ⇑ aux structures
    - chaînes de référence
    - CR de haut et de bas niveau résultant de la mise en relation des maillons relevant du même référent
  - ⇑ des indices
    - maillons
    - expressions référentielles
- cf. projet DemoCrat [Landragin, 2015]



# Adapter \_\_\_\_\_me au projet



## Conserver

- l'idée (et l'expertise) d'un prémarquage et d'une annotation assistée
- l'annotation de chaînes "topicales"

## Oublier

- l'annotation "descendante" et naïve

## Proposer

- une annotation "ascendante" et experte (identifier les maillons)
  - mais annotation exhaustive peu envisageable (chronophage ++ face aux textes non standards)
- ⇒ annotation des maillons liés à la consigne d'écriture  
(*"continue/raconte l'histoire de"*, ... la consigne ResolCo)

# Plan

- 1 Projets ANR AnnoDis et E :Calm
- 2 Collecte des données : la consigne ResolCo**
- 3 Annotation des continuités référentielles dans les données RésolCo
- 4 Résultats attendus et à venir



## Une seule consigne, identique du CE2 au Master

*Racontez une histoire dans laquelle vous insérerez, séparément et dans l'ordre donné, les trois phrases suivantes :*

Elle habitait dans cette maison depuis longtemps.

Il se retourna en entendant ce grand bruit.

Depuis cette aventure, les enfants ne sortent plus la nuit.

*Vous pouvez découper les bandelettes contenant les phrases ci-dessous ou bien recopier chaque phrase avec soin à l'identique de celles qui vous sont données.*

## Une consigne conçue pour provoquer des stratégies pour gérer la cohérence

Une tâche-problème de cohérence [Garcia-Debanc, 2016] : *Elle ? cette maison ? Il ? ce grand bruit ? les enfants ? quelle aventure ? la nuit ? etc.*



# Une consigne provoquant des problèmes de cohérence

01/02/2019

qui arrive-t-il  
aux enfants?  
Que fait l'ogresse?

ogresse

Il était une fois une petite fille qui s'appelle Baba Yaga elle habitait dans cette maison antoi depuis longtemps. Deux petit enfan son parti lui rendre visite Lola et Max. Il se ~~sont~~ retourna en entendant ce grand bruit d'ogresse ensuite les enfants ont couru et Depuis cette aventure les enfants ne sortent plus la nuit.

copie de CM1

# Une consigne provoquant des problèmes de cohérence

ogresse

Étais une fois une petite fille qui s'appelle Baba Yaga elle habitait dans cette maison antoi depuis longtemps.

Deux petit enfant son parti lui rendre visite Lola et Max. Il se ~~sont~~ retourna en entendant ce grand bruit d'ogresse ensuite les enfants ont couru ~~et~~ Depuis cette aventure les enfants ne sortent plus la nuit.

01/02/2019

qui arrive-t-il aux enfants?

Que fait l'ogresse?

copie de CM1

- Quelles stratégies pour introduire les référents acteurs (*Elle, Il, les enfants*) ?
- Quelles stratégies pour maintenir la continuité référentielle ?
- Comment gérer la cohabitation entre les chaînes de référence ?

# Plan

- 1 Projets ANR AnnoDis et E :Calm
- 2 Collecte des données : la consigne ResolCo
- 3 Annotation des continuités référentielles dans les données RésolCo
  - Des données primaires aux données annotables
  - Une annotation sélective et adaptée
- 4 Résultats attendus et à venir

# Des données primaires aux données annotables

## Transcription et encodage des traces d'écriture et interventions

```

<?xml version='1.0' encoding='UTF-8' ?>
<TEI>
<teiHeader>
  <front>
    <p>Nom de l'élève</p>
    <p>01/02/2019</p>
  </front>
  <body>
    <p>
      Il étai une foix <mod type="subst"><del>une petite fiell</del><add>
      ogrèse</add></mod> qui s'apelle BabaYaga <lb/>elle habitait dans cette
      maison antai depuis longtems. <lb/>Deux petit enfan son parti lui
      rendre visite Lola et <lb/>Max. Il se <mod type="subst"><del>ont
      </del></mod> retourna en entendant ce grand bruit <lb/>d'ogrèse ensuite
      les enfants ont couru et Depuis cette <lb/>aventure les enfants ne
      sortent plus la nuit.
    </p>
  </body>
</back>
  <metamark who="teacher">Qu'arrive-t-il aux enfants ?</metamark>
  <metamark who="teacher">Que fait l'ogresse ?</metamark>

```

ogresse

Il étai une foix une petite fiell qui s'apelle Baba Yaga elle habitait dans cette maison antai depuis longtems.

01/02/2019

Deux petit enfan son parti lui rendre visite Lola et Max. Il se ~~ont~~ retourna en entendant ce grand bruit d'ogrèse ensuite les enfants ont couru et Depuis cette aventure les enfants ne sortent plus la nuit.

qui arrive-t-il aux enfants?

Que fait l'ogresse?

# Des données primaires aux données annotables

## Normalisation de l'orthographe

- Annoter de l'écrit standard vs. non normé
- Détection des erreurs via Glozz selon le guide de normalisation orthographique (ScolEdit, Ecriscol)
- Possibilité de proposer plusieurs corrections
- Permet d'appliquer des modules de TAL (analyse syntaxique)

*Il étai une foix une ogrése qui s'apelle...*

→ *Il était une fois une ogresse qui s'appelle...*

Aperçu provisoire [Ho-Dac et al., 2020]

niveau	#textes	#mots	% err
CE2	31	3252	12.8
CM1	37	4823	13.7
CM2	42	8351	12.8
6e	45	7860	13.1
4e	15	4887	12.6
3e	27	8622	9.2
Master	12	5318	2.1



# Annotation des continuités référentielles dans RésolCo

Commencer par les continuités "provoquées" par la consigne et référant aux personnages

Camille et Marie jouaient la nuit dans la forêt.

Pout à coup elle eut une idée : « Viens Marie je t'ai  
vrais te montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps.

Et dans la maison

Il se retourna en entendant ce grand bruit.

Elle monta dans

la grenier et ils entendirent un autre bruit. Et ils vi-  
rent une ombre de rat géant. Et ils retournèrent  
chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

copie de CM1



# Annotation des continuités référentielles dans RésolCo

Considérer une continuité entre : [Camille, elle], [je, ma], [elle], [ils+, eux, les enfants]



# Annotation de la co-référence dans les données RésolCo

- ① Délimiter les maillons "Elle", puis les maillons "Il", puis les maillons "les enfants"
  - S'aider d'un prémarquage et utiliser Glozz (cf. ANNODIS [Péry-Woodley et al., 2011]) :
    - pronoms 3e pers. suj et obj
    - noms propres
    - déterminants possessifs de 3e pers.
    - SN *le sien, le leur, les enfants*
  - Utilisation du guide DemoCrat pour la délimitation des expressions référentielles [Lattice et al., 2019]
  - Considérer les passages au discours direct comme une co-référence
  - Considérer les intégrations d'un référent dans un collectif comme une co-référence (le trait "groupe")
  - Annoter deux incertitudes : rattachement et délimitation
- ② Relier les maillons "Elle", "Il" et "les enfants" pour vérifier l'annotation

# Annotation de copies normalisées et prémarquées

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Features Informations

Feature name	Feature value
type	

Color mode: STYLESHEET

Unit style	Relation style	Schema style
Type name	Background-color	Hide
maillon-candidat		<input type="checkbox"/>
maillon Elle		<input checked="" type="checkbox"/>
maillon Il		<input checked="" type="checkbox"/>
maillon lesEnfants		<input checked="" type="checkbox"/>
phraseConsigne		<input type="checkbox"/>

Model As Text Tools

Annotation model:

Units	Relations	Schemas
maillon-candidat	coref Elle	
maillon Elle	coref Il	
maillon Il	coref lesEnfants	
maillon lesEnfants		
phraseConsigne		

# Délimitation des maillons\_Elle

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations	
Feature name		Feature value	
type		maillon_Elle	
groupe		Non	
incertitude sur la délimitation		Non	
incertitude sur le rattachement		Non	
commentaire			

Color mode: **STYLESHEET**

Unit style		Relation style		Schema style	
Type name	Background-color		Hide		
maillon-candidat			<input type="checkbox"/>		
maillon_Elle			<input type="checkbox"/>		
maillon_Il			<input type="checkbox"/>		
maillon_lesEnfants			<input checked="" type="checkbox"/>		
phraseConsigne			<input checked="" type="checkbox"/>		

Model	As Text	Tools
<b>Annotation model:</b>		
<b>Units</b>	<b>Relations</b>	<b>Schemas</b>
maillon-candidat	coref_Elle	
maillon_Elle	coref_Il	
maillon_Il	coref_lesEnfants	
maillon_lesEnfants		

# Délimitation des maillons\_II

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations	
		Feature name	Feature value
type			maillon_II
groupe			Non
incertitude sur la délimitation			Non
incertitude sur le rattachement			Non
commentaire			

Color mode: STYLESHEET

Unit style	Relation style	Schema style	
Type name	Background-color		Hide
maillon-candidat			<input checked="" type="checkbox"/>
maillon_Elle			<input type="checkbox"/>
maillon_II			<input type="checkbox"/>
maillon_lesEnfants			<input checked="" type="checkbox"/>
phraseConsigne			<input type="checkbox"/>

Model As Text Tools

Annotation model:

Units	Relations	Schemas
maillon-candidat	coref_Elle	
maillon_Elle	coref_II	
maillon_II	coref_lesEnfants	
maillon_lesEnfants		

# Délimitation des maillons\_lesEnfants

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins SandBox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ». Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations	
		Feature name	Feature value
type		maillon_lesEnfants	
groupe		Non	
incertitude sur la délimitation		Non	
incertitude sur le rattachement		Non	
commentaire			

Color mode: STYLESHEET

Unit style	Relation style	Schema style	
Type name	Background-color		Hide
maillon-candidat			<input type="checkbox"/>
maillon Elle			<input type="checkbox"/>
maillon Il			<input checked="" type="checkbox"/>
maillon_lesEnfants			<input type="checkbox"/>
phraseConsigne			<input type="checkbox"/>

Model	As Text	Tools
<b>Annotation model:</b>		
<b>Units</b>	<b>Relations</b>	<b>Schemas</b>
maillon-candidat	coref Elle	
maillon Elle	coref Il	
maillon Il	coref_lesEnfants	
maillon_lesEnfants		



# Construction de la CR\_Elle

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins SandBox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ». Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant. Depuis cette aventure, les enfants ne sortent plus la nuit.

Features Informations

Feature name	Feature value
type	maillon_lesEnfants
groupe	Non
incertitude sur la délimitation	Non
incertitude sur le rattachement	Non
commentaire	

Color mode: STYLESHEET

Unit style Relation style Schema style

Type name	Line-color	Hide
coref_Elle		<input type="checkbox"/>
coref_II		<input checked="" type="checkbox"/>
coref_lesEnfants		<input checked="" type="checkbox"/>

Model As Text Tools

Annotation model:

Units	Relations	Schemas
maillon-candidat	coref_Elle	
maillon_Elle	coref_II	
maillon_II	coref_lesEnfants	
maillon_lesEnfants		

# Construction de la CR\_II

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins SandBox ?

Camille et **Manu** jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « **Viens, Manu, je vais te** montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps. Et dans la maison **il se retourna en entendant ce grand bruit. Ils** montèrent dans le grenier et **ils entendirent un autre bruit. Et ils** virent une ombre de rat géant. Et **ils retournèrent chez eux** en courant.

Depuis cette aventure, **les enfants** ne sortent plus la nuit.

**Features Informations**

Feature name	Feature value
type	maillon_lesEnfants
groupe	Non
incertitude sur la délimitation	Non
incertitude sur le rattachement	Non
commentaire	

Color mode: **STYLESHEET**

**Unit style Relation style Schema style**

Type name	Line-color	Hide
coref Elle	Blue	<input checked="" type="checkbox"/>
coref Il	Green	<input type="checkbox"/>
coref lesEnfants	Orange	<input checked="" type="checkbox"/>

**Model As Text Tools**

**Annotation model:**

Units	Relations	Schemas
maillon-candidat	coref Elle	
maillon Elle	coref Il	
maillon il	coref lesEnfants	
maillon lesEnfants		
phraseConsigne		

# Construction de la CR\_lesEnfants

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ».

Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Features Informations

Feature name	Feature value
type	maillon_lesEnfants
groupe	Non
incertitude sur la délimitation	Non
incertitude sur le rattachement	Non
commentaire	

Color mode: STYLESHEET

Unit style Relation style Schema style

Type name	Background-color	Hide
maillon-candidat		<input checked="" type="checkbox"/>
maillon_Elle		<input checked="" type="checkbox"/>
maillon_Il		<input checked="" type="checkbox"/>
maillon_lesEnfants		<input checked="" type="checkbox"/>
phraseConsigne		<input checked="" type="checkbox"/>

Model As Text Tools

Annotation model:

Units	Relations	Schemas
maillon-candidat	coref_Elle	
maillon_Elle	coref_Il	
maillon_Il	coref_lesEnfants	
maillon_lesEnfants		
phraseConsigne		

# Annotation complète de la copie

Glozz - 2.1 - Logged as hodaclm / EC-CM1-2015-TFGLX-D1-R12-V1\_T\_norm\_realigned.ac

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Camille et Manu jouaient la nuit dans la forêt. Tout à coup elle eut une idée : « Viens Manu je vais te montrer ma cabane ». Elle habitait dans cette maison depuis longtemps. Et dans la maison il se retourna en entendant ce grand bruit. Ils montèrent dans le grenier et ils entendirent un autre bruit. Et ils virent une ombre de rat géant. Et ils retournèrent chez eux en courant. Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations
Feature name		Feature value
type		maillon_lesEnfants
groupe		Non
incertitude sur la délimitation		Non
incertitude sur le rattachement		Non
commentaire		

Color mode: STYLESHEET

Unit style		Relation style	Schema style
Type name	Line-color		Hide
coref_Elle	Blue		<input type="checkbox"/>
coref_Il	Green		<input type="checkbox"/>
coref_lesEnfants	Orange		<input type="checkbox"/>

Model As Text Tools

Annotation model:

Units	Relations	Schemas
maillon-candidat	coref_Elle	
maillon_Elle	coref_Il	
maillon_Il	coref_lesEnfants	
maillon_lesEnfants		
phraseConsigne		

# Enjeux de l'annotation d'écrits non standards

## Une annotation adaptée

- Prévoir une annotation permettant de repérer les problèmes de cohérence sans les résoudre ( éviter une lecture "indulgente")

Ex *Paul et sa soeur, Pauline aimaient se disputer. [II]\_ ? se retourna en entendant ce grand bruit.*

**[II]\_ ?** : Selon la suite de l'histoire, peut-être rattaché à Paul avec incertitude de rattachement ou n'être rattaché à rien

⇒ Un **guide d'annotation** le plus précis possible avec auto-évaluation

# Plan

- 1 Projets ANR AnnoDis et E :Calm
- 2 Collecte des données : la consigne ResolCo
- 3 Annotation des continuités référentielles dans les données RésolCo
- 4 Résultats attendus et à venir

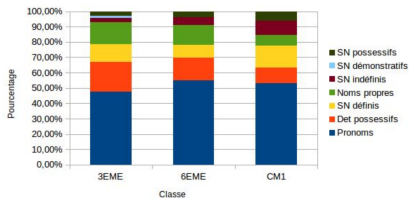
# Un gros chantier qui a eu du mal à démarrer ...

- Une sous-estimation du travail de préparation des données (2 ans de transcription et normalisation)
- Une sous-estimation de la réflexion nécessaire pour adapter
  - la méthode et les outils d'annotation : unités ? traits ? schémas ? relations ?
  - les modèles d'annotation existants : quelles expressions référentielles et relations de co-référence considérer ?
- Un premier travail d'annotation des CR\_Elle et des résultats encourageants

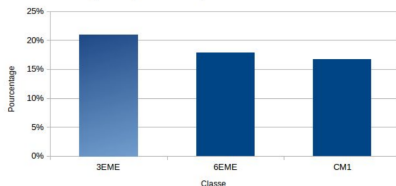
# Premiers résultats concernant les CR\_Elle

Roudaut, L. (2020) Les traces de continuité référentielle dans les écrits scolaires, mémoire de 1ere année de Master LITL, UT2J.

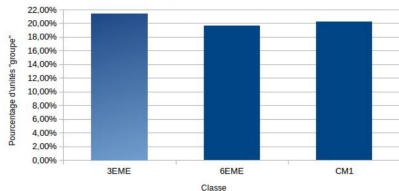
Répartition des natures des maillons en fonction des niveaux scolaires



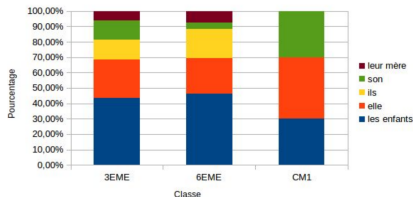
Pourcentage de copies dans lesquelles le 1er maillon est le "Elle"



Pourcentage d'unités "groupe" selon les niveaux scolaires



Formes des derniers maillons en fonction des niveaux scolaires





... qui a eu du mal à démarrer, mais un chantier lancé

### Des décisions qui semblent efficaces

- Choisir les personnages principaux comme point de départ (adaptation à d'autres consignes d'écriture)
- Annoter en autant d'étapes que de référents
- Annoter "fois deux" en reliant les maillons
- Le trait "groupe"

### Un projet encore plus ambitieux

... qui a eu du mal à démarrer, mais un chantier lancé

## Des décisions qui semblent efficaces

### Un projet encore plus ambitieux

- Tester l'annotation sur d'autres corpus :
  - ScolEdit et ÉNEPS, (cf. [Pallantin et Jacques](#))
- Analyser les annotations récoltées
  - comparer les continuités avec celles observées dans des écrits normés (cf. [Federzoni](#))
  - définir des indicateurs de compétences (cf. [Garcia-Debanc](#))
- D'autres annotations de structures discursives en cours
  - Segmentation : unités de discours élémentaires, unités syntaxiques, unités graphiques
  - Annotation des relations rhétoriques (ANNODIS\_rr)
- À croiser avec les traces d'écriture, les normalisations et les jugements d'enseignant.e.s

C'était une sombre soirée d'octobre, plus précisément, le soir d'Halloween. Une bande d'adolescents composée de 2 filles et 2 garçons avait décidé de faire le tour du quartier pour réclamer des bonbons aux résidents.

Après avoir passé le coin le plus lumineux de la ville, ils se dirigèrent vers celui qui faisait trembler les plus petits. Ana et Alice deux des 4 membres de la petite troupe, ayant trop peur de s'aventurer dans ce quartier sombre, décidèrent d'attendre leurs amis, Hugo et Swan au début de la rue, là où se trouvaient encore quelques rares lampadaires.

Les deux jeunes hommes se dirigèrent alors vers la seule maison, se trouvant au bout de cette longue et interminable rue. Cette maison ressemblant à un manoir, était à peine distincte du ciel noir. Une vieille femme y vivait. Elle habitait dans cette maison depuis longtemps. Ils toquèrent, mais ne reçurent aucune réponse. Ils répétèrent ce geste à plusieurs reprises, mais c'était toujours le même résultat. Alors que Swan commença à s'impatienter, Hugo entreprit de toquer une dernière fois. Ils s'attendirent à ne recevoir aucune réponse, mais c'est alors que la porte s'ouvrit doucement dans un long grincement. Et alors que Swan posa un pied dans la maison, toutes les bougies se trouvant dans le couloir s'allumèrent. Sans une once de peur, les deux jeunes hommes, croyant à une blague, s'aventurèrent dans ce long couloir. Alors que Swan voulait dire à son ami " que cette baraque était trop stylée ", il remarqua qu'Hugo avait disparu. Swan de moins en moins confiant, se décida à continuer à avancer. Quand tout à coup... il se retourna en entendant ce grand bruit. Et avant même d'avoir pu pousser ne serait-ce qu'un cri de peur, ses longs ongles transpercèrent son cou, tuant ce jeune homme sur le coup.

Quelques heures plus tard, ne voyant toujours pas leurs amis rentrer, Ana et Alice appelèrent la police. Au petit matin, sur le journal matinal, nous pouvions voir en gros titre " deux jeunes garçons retrouvés décapité pour l'un et pendu pour l'autre, dans la maison de Sophia Smith décédée depuis plus de 4 ans ".

Depuis cette aventure, les enfants ne sortent plus la nuit. Surtout celle d'Halloween.



Doquet, C., David, J., Fleury, S., and (Eds) (2017).

Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement.

In *Corpus [Online]*, volume 16 (Special Issue). OpenEdition.



Garcia-Debanc, C. (2016).

Une tâche problème pour analyser les compétences d'élèves de sixième en matière de cohésion textuelle.

In L., S., D., V., and (coord)., C. B., editors, *Connexion et indexation. Ces liens qui tissent le texte*, Connexion et indexation. Ces liens qui tissent le texte, pages 263–278. ENS Editions.



Garcia-Debanc, C., Ho-Dac, L.-M., Bras, M., and Rebeyrolle, J. (2017).

Vers l'annotation discursive de textes d'élèves.

*Corpus*, 16.



Ho-Dac, L.-M., Fleury, S., and Ponton, C. (2020).

E : Calm resource : a resource for studying texts produced by french pupils and students.

In *LREC*.



Landragin, F. (2015).

Description, modélisation et détection automatique des chaînes de référence (democrat).

*Bulletin de l'AFIA*, 92 :11–15.



Lattice, LiLPa, ICAR, and IHRIM (2019).

Democrat.

ORTOLANG (Open Resources and TOols for LANGuage) –[www.ortolang.fr](http://www.ortolang.fr).



Péry-Woodley, M.-P., Afantenos, S. D., Ho-Dac, L.-M., and Asher, N. (2011).

La ressource annodis, un corpus enrichi d'annotations discursives.

*Traitement Automatique des Langues*, 52(3) :71–101.



Péry-Woodley, M.-P., Ho-Dac, L.-M., Rebeyrolle, J., Tanguy, L., and Fabre, C. (2017).

A corpus-driven approach to discourse organisation : from cues to complex markers.

*Dialogue and Discourse*, 8(1) :66–105.