



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

**Jeudi 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 du 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><br>Lydia-Mai HO-DAC, Silvia FEDERZONI, Josette REBEYROLLE, Claudine GARCIA-DEBANC<br><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><br>Luca PALLANTI & Marie-Paule JACQUES<br><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><br>Silvia FEDERZONI<br><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><br>Claudine GARCIA-DEBANC<br><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><br>Frédéric LANDRAGIN<br><i>Laboratoire LATTICE, ENS, UMR 8094</i>   |

