Silvia Federzoni

Doctorante (CLLE) & IE (LLL)

Université Toulouse Jean Jaurès (CLLE, UMR 5263)
Université d'Orléans (LLL, UMR 7270)
silvia.federzoni@univ-orleans.fr

Formation

Depuis octobre 2019 **Doctorat en Sciences du Langage**, *Université Toulouse Jean Jaurès*, CLLE (Cognition, Langues, Langage, Ergonomie - UMR 5263), Contrat Doctoral Unique - Prolongation de 3 mois obtenue suite au COVID-19 - Soutenance prévue pour juillet 2024

Titre *Typologie des chaînes de référence à la lumière de corpus annotés diversifiés*

Directrices Cécile Fabre & Lydia-Mai Ho-Dac

Dans le cadre de ma thèse je m'intéresse aux chaînes de (co)référence et à leur fonctionnement textuel. J'aborde cette question en prenant en compte différents domaines : celui de la sémantique référentielle, de l'analyse du discours et du traitement automatique des langues. L'objectif de mon travail de recherche est de proposer une typologie de ces structures discursives à partir de l'analyse de corpus diversifiés annotés en chaînes de référence. Pour ce faire, je propose une nouvelle approche consistant à analyser les chaînes de référence dans leur linéarité. Cela consiste à s'intéresser aux modalités d'enchaînement des expressions référentielles dans les textes, ce qui permet d'aller au-delà de la description des caractéristiques globales des chaînes, très largement proposée dans les travaux existants. Afin d'analyser de manière systématique les chaînes, tout en restant dans la linéarité du discours, je propose une méthode d'analyse outillée de corpus. La méthode que je propose consiste dans la combinaison des techniques du Traitement Automatique des Langues et des techniques d'analyse des séquences, utilisées traditionnellement en Sciences Sociales. Les analyses menées jusqu'à présent m'ont permis d'obtenir des résultats intéressants, qui confirment également la validité de la méthode.
La validité de la méthode et son intérêt ont été confirmés également par des chercheurs bien reconnus dans le domaine lors des présentations orales que j'ai faites dans le cadre de colloques internationaux (Federzoni et al., 2021). J'ai également présenté les premiers résultats concernant l'impact de la nature du référent (humain ou non humain) sur les modalités d'enchaînement des expressions référentielles dans des colloques internationaux et nationaux (Federzoni, 2020 et Federzoni et al., 2021). J'ai publié un chapitre dans un ouvrage collectif présentant en détail la nouvelle approche que je propose, la méthode et les résultats obtenus (Federzoni et al., 2023). L'objectif était de poser clairement la problématique et l'approche consistant à rester dans la linéarité afin d'avoir des retours constructifs de la part de chercheurs de haut niveau et de faire connaître mon travail de recherche.

2018–2019 **Master 2 en Sciences du Langage**, *Parcours Linguistique, Informatique et Technologies du Langage (LITL)*, Université Toulouse Jean Jaurès, mention TB

Mémoire : *Évaluation et exploitation de la Ressource ANNODIS pour la détection des chaînes de référence*, dir. Lydia-Mai Ho-Dac & Josette Rebeyrolle

2017–2018 **Master 1 en Sciences du Langage**, *Parcours Linguistique, Informatique et Technologies du Langage (LITL)*, Université Toulouse Jean Jaurès, mention B

Mémoire : *Étude de la présence et du fonctionnement des marqueurs de relations conceptuelles dans un corpus aligné franco-italien*, dir. Anne Condamines & Cécile Fabre

2013-2017 **Master en Langues et littératures modernes, européennes et américaines**, *Parcours Sciences linguistiques, littéraires et de la traduction*, Université de Rome La Sapienza, mention TB (105/110)

Mémoire : *Les marqueurs de la relation d'hyperonymie en français et en italien*, dir.Marie Pierre Escoubas-Benveniste & Isabella Chiari

2008-2012 **Licence en Philosophie, lettres, sciences humaines et études orientales**, *Parcours Médiation linguistique et culturelle* , Université de Rome La Sapienza, mention B (100/110)

Mémoire : *Expressions figées tirées du langage du sport*, dir. Daria Galateria

2008 **Baccalauréat**, *Études commerciales en langues étrangères* , ITCG Guido Baccelli, Civitavecchia (Rome), mention TB (100/100)

Expériences d'enseignement

J'ai dispensé des enseignements variés à différents niveaux de la formation des étudiants au département de Sciences du Langage à l'Université Toulouse Jean Jaurès. Pendant ces années j'ai assuré les tâches liées à la gestion des TD, des ressources en ligne sur la plateforme ENT (Moodle) et des contenus d'examens.

2022-2023 **Chargée de cours**, *Université Toulouse Jean Jaurès* , UFR LLCE - Département Sciences du Langage

2020-2022 **Enseignante contractuelle (avenant d'enseignement CDU)**, *Université Toulouse Jean Jaurès*, UFR LLCE - Département Sciences du Langage

Enseignements dispensés

Licence 1 **SL00105T - Découvrir les SDL et outils d'analyse en SDL 1**, TD - 24h

Le cours a pour objectif d'initier les étudiants à la recherche documentaire et à la manipulation de données langagières en Sciences du Langage. Dans une première partie du cours, les étudiants découvrent et interrogent les outils offerts par les centres de ressources et les bibliothèques universitaires afin de produire un dossier documentaire sur une thématique en lien avec les Sciences du Langage. Dans une deuxième partie du cours, les étudiants découvrent, explorent et manipulent des données langagières, notamment des corpus et des bases de données. L'objectif de cette partie du cours est d'accompagner les étudiants dans une démarche scientifique leur permettant de confronter des hypothèses sur le langage aux usages réels. Les étudiants suivent des séances pratiques sur ordinateurs pendant lesquelles ils interrogent des corpus et analysent les résultats obtenus. Les deux parties du cours, alternent séances en présentiel et travail en autonomie.

SL00202V - Histoire de la linguistique, CM/TD - 24h

Le cours a pour objectif de faire connaître les différents courants de la linguistique, à partir de Port-Royal et la grammaire générale jusqu'à la grammaire générative de Chomsky. J'ai assuré le cours à distance et en hybride pendant la période de confinement, en proposant des activités adaptées à la situation. Pour augmenter l'implication des étudiants à distance j'ai mis en place des révisions de cours, effectuées régulièrement via des questionnaires interactifs en exploitant des plateformes comme Kahoot ou Wooclap.

Licence 2 **SL00404T - Discours oral et écrit**, CM/TD - 12h

Ce cours propose une étude des constructions typiques de l'oral, telles les dislocations ou les clivées. En s'appuyant sur une analyse des différences entre l'écrit normé et l'oral spontané, le cours mène les étudiants à analyser les facteurs discursifs qui peuvent influencer la manière dont on s'exprime (canal de communication, situation d'énonciation, etc.). Pendant le cours, les étudiants sont confrontés régulièrement à l'analyse des données réelles leur permettant de contraster l'oral spontané à l'écrit normé, d'identifier les constructions typiques de l'oral, notamment les dislocations et les clivées, et de réfléchir à leur fonctions.

SL00405T - Outils d'analyse en SDL 4, TD - 18h

Le cours est divisé en deux parties : l'une porte sur la présentation de étapes à suivre pour une collecte des données en Sciences du Langage ; l'autre porte sur la formalisation du langage et la manipulation de textes numériques. J'ai assure la deuxième partie du cours. Les étudiants suivent des séances pratiques sur ordinateurs pendant lesquelles ils effectuent des recherches et des transformations des chaînes de caractères en utilisant des automates à états finis et des expressions régulières.

PE00304T - Discours oral / discours écrit (Mathématiques 1 / Français 1), *UFR SES - Département Mathématiques, Informatique*, CM/TD - 12h

Ce cours s'adresse aux étudiants inscrits à la mineure *Professorat des écoles*. Il est divisé en deux parties. J'ai assuré la première partie intitulée *Discours oral / Discours écrit*. Cette partie du cours porte sur la comparaison des productions typiques de l'oral spontané à celles typiques de l'écrit normé. L'objectif est d'amener les étudiants, futurs professeurs d'école, à examiner certains aspects de l'apprentissage de l'écrit chez les enfants. Pendant le cours, les étudiants sont confrontés régulièrement à l'analyse des données réelles leur permettant de contraster l'oral spontané à l'écrit normé et d'identifier les constructions typiques de l'oral, notamment les dislocations et les clivées. Ce cours se différencie du cours SL00404T par le fait que l'accent est mis sur les aspects de l'apprentissage de l'écrit chez les enfants. Les fonctions discursives des constructions étudiées ne sont pas abordées en détail dans ce cours.

Licence 3 **SL00504T - Histoire des idées linguistiques**, CM/TD - 2h

Ce cours a pour objectif de faire connaître aux étudiants les projets de recherche menés par les différents membres de l'équipe pédagogique. Chaque semaine, un ou deux membres de l'équipe pédagogique présentent leurs recherches. Dans ce contexte, j'ai proposé une présentation de mon projet de thèse, en explicitant le cadre théorique et la méthodologie utilisée. J'ai aussi présenté les premiers résultats de ma recherche.

SL00504V - Analyse de corpus, TD - 25h

Le cours a pour objectif de faire découvrir aux étudiants l'analyse outillée des données langagières. Les étudiants apprennent à constituer un corpus et ensuite à le manipuler via des outils conçus pour l'exploration des corpus. Le TD que j'ai assuré était basé sur la manipulation de l'outil AntConc. L'objectif était d'amener les étudiants à effectuer des recherches sur corpus et analyser les résultats obtenus. J'ai assuré le cours à distance pendant la période de confinement.

Master 2 **SLT0113T - Thématiques actuelles de la recherche en TAL**, TD - 22h

Ce cours propose une série de séminaires et de tutoriels dans le domaine du Traitement Automatique des Langues. Dans le cadre de cette UE, j'ai assuré :

— 3 séminaires (4h)
— 3 tutoriels (18h)

Chaque année j'ai donné un séminaire portant sur mon travail de thèse. J'ai mis l'accent sur les questions théoriques et les choix méthodologiques effectués. À chaque fois, le séminaire a servi d'introduction au tutoriel que j'ai conçu pour cette même UE. J'ai proposé un tutoriel sur la classification automatique des chaînes de référence qui a permis aux étudiants de master de découvrir et d'appliquer des techniques de clustering hiérarchique. Chaque année, j'ai adapté le tutoriel en fonction des cours que les étudiants avaient eu pendant l'année. Les objectifs étaient de leur faire manipuler des fichiers annotés dans des formats variés (xml, csv), de leur faire utiliser des méthodes de clustering hiérarchique (utilisant python ou R) et de les amener à analyser les résultats d'un point de vue linguistique avec un retour aux données. Pour ce faire, je leur ai demandé de travailler sur des corpus du français annotés en chaînes de référence.

Niveau	Matière	Code UE	Type	Heures	Période
Licence 1	Découvrire les SDL et outils d'analyse en SDL 1	SL00105T	TD	24h	2021-2022
Licence 1	Histoire de la linguistique	SL00202V	CM/TD	24h	2020-2021
Licence 2	Discours oral et écrit	SL00404T	CM/TD	12h	2022-2023
Licence 2	Outils d'analyse en SDL 4	SL00405T	TD	18h	2021-2022
Licence 2	Mathématiques 1 / Français 1	PE00304	CM/TD	12h	2021-2022
Licence 3	Histoire des idées linguistiques	SL00504T	CM/TD	2h	2021-2022
Licence 3	Analyse de corpus	SL00504V	TD	25h	2020-2021
Master 2	Thématiques actuelles de la recherche en TAL	SLT0113T	TD	22h	2020-2023
Total				**139h**	

Encadrement de mémoires

Master 1 **Sciences du Langage**

Pendant mon avenant d'enseignement j'ai également co-encadré cinq mémoires de master 1 en Sciences du Langage, parcours Linguistique, Informatique et Technologies du Langage (LITL).

En lien avec mon sujet de thèse :

Degrutère, Solène (2022). *Transitions syntaxiques au fil des chaînes de référence selon le genre textuel*, mémoire Master 1 LITL, co-encadré avec Cécile Fabre, Université Toulouse Jean Jaurès

Villedey, Judith (2022). *L'identification des chaînes topicales au sein d'un corpus annoté en chaînes de référence*, mémoire de Master 1 LITL, co-encadré avec Lydia-Mai Ho-Dac, Université Toulouse Jean Jaurès

Mayer, Mélanie (2021). *Les caractéristiques sémantiques de l'expression référentielle : valeurs spécifique et générique dans les chaînes de référence*, mémoire de Master 1 LITL, co-encadré avec Cécile Fabre, Université Toulouse Jean Jaurès

Roudaut, Laura (2020). *Les traces de continuité référentielle dans les écrits scolaires*, mémoire de Master 1 LITL, co-encadré avec Lydia-Mai Ho-Dac, Université Toulouse Jean Jaurès

En lien avec le cours *Discours oral et écrit* :

Lidén, Maria (2022). *Évaluation de la place des constructions typiques de l'immédiat dans les écrits scolaires. Le cas des constructions clivées*, mémoire de Master 1 LITL, co-encadré avec Josette Rebeyrolle, Université Toulouse Jean Jaurès

Autres expériences

Depuis octobre 2023 **IE projet DOING**, *Laboratoire Ligérien de Linguistique (LLL - CNRS & Université d'Orléans)*
Extraction d'informations : resolution de coréférences et extraction de relations temporelles dans le domaine médical

2018 – 2019 **IE projet E-Calm**, *CLLE (CNRS & UT2J)*, Université Toulouse Jean Jaurès
En tant qu'ingénieure d'étude j'étais responsable de la constitution et de l'annotation d'un corpus d'écrits scolaires (RésolCo : corpus dédié à la Résolution de problèmes de cohésion textuelle).

Tâches réalisées :

- Mise en place de campagnes de transcription et d'annotation de données au format XML, suivant la norme TEI
- Rédaction de guides de bonnes pratiques pour la transcription des copies d'élèves et leur annotation
- Gestion et formation des stagiaires (niveaux L3, M1 et M2)
- Présentation des avancées auprès des responsables du projet
- Proposition d'un site web pour la présentation, l'exploration et la diffusion du corpus

2017 **Stage Erasmus+ Traineeship**, *CLLE (CNRS & UT2J)*, Université Toulouse Jean Jaurès
Ce stage portait sur l'étude de la relation d'hyperonymie dans un corpus aligné français-italien. Il m'a permis de découvrir le monde de la recherche, de consolider mes compétences et mes connaissances en traduction, et de m'engager progressivement dans une approche de linguistique de corpus outillée.

Tâches réalisées :

- Prise en main d'outils pour l'analyse de cropus : AntConc, Alinéa, TXM
- Alignement de corpus parallèles (fr-it)
- Analyse semi-automatique de corpus (repérage des marqueurs de la relation d'hyperonymie et analyse des phénomènes rencontrés)
- constitution d'un corpus italien en langue de spécialité
- participation aux séminaires

Formations complémentaires

10-11 septembre 2021 **TEI 2 : Encoder en XML-TEI (niveau avancé)**, *Centre d'études supérieures de la Renaissance*, Tours
Cette formation m'a permis de consolider mes connaissances dans les bonnes pratiques pour le balisage des corpus au format XML suivant la norme TEI. J'ai appris comment structurer les données et à gérer les différentes couches d'annotation présentes dans les corpus annotés en structures discursive que j'ai l'habitude de manipuler.
Équipe pédagogique : Lou Burnard, Mathieu Duboc & Elena Pierrazzo

2017 **École d'été ESSLLI**, *Université de Toulouse 1 Capitole*, European Summer School in Logic, Language and Information.
Pendant cette école d'été j'ai pu suivre des cours en linguistique, logique et traitement automatiques des langues. Ces cours sont dispensés par des chercheurs de haut niveau et se basent sur un apprentissage pratique.

Responsabilités

2020-2022 **Représentante des doctorant.e.s du laboratoire CLLE**, Université Toulouse Jean Jaurès

- Organisation de la journée d'étude « Printemps des Jeunes Chercheurs CLLE », Université Toulouse Jean Jaurès
- Organisation de la Journée de Rentrée des Doctorants CLLE, Université Toulouse Jean Jaurès

2019-2022 **Co-responsable du collectif CEPEL (Cercle Étudiant Pour l'Étude du Langage)**, Université Toulouse Jean Jaurès

- Organisation de séminaires de vulgarisation scientifique, Université Toulouse Jean Jaurès
- Gestion du site web : présentation du collectif et de ses membres, présentation des séminaires

Publications

Chapitres dans un ouvrage collectif

Federzoni, S., Ho-Dac, L.-M. et Fabre, C. (2023). *A linear approach of chain composition*. In Gardelle, L., Vincent-Durroux, L. & Vinchel-Rosin, H. (dir.), *Reference : from Convention to Pragmatics*, Amsterdam/Philadelphia , John Benjamins. pp. 107-126.

Condamines, A., Escoubas Benveniste, MP. et Federzoni, S. (2021). *Apport de la traduction dans l'étude des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien*. In Frérot, C., Pecman, M. *Des corpus numériques à la modélisation linguistique en langues de spécialité.* Grenoble : Presses de l'UGA. pp.313-336.

Articles dans une revue

Condamines, A., Escoubas Benveniste, MP. et Federzoni, S. (2022). Apport d'un corpus de presse spécialisé parallèle français/italien à l'analyse des marqueurs et de la relation de méronymie. *Éla. Études de linguistique appliquée*, 208, pp. 429-446.

Publications dans des actes de colloques internationaux

Federzoni, S., Ho-Dac, L.-M. et Fabre, C. (2021). *Coreference Chains Categorization by Sequence Clustering*. Dans Association for Computational Linguistics (dir.), 2nd Workshop on Computational Approaches to Discourse. Punta Cana, République Dominicaine. pp. 52-57

Federzoni, S., Ho-Dac, L.-M. et Rebeyrolle, J. (2020). *Les chaînes topicales dans la ressource ANNODIS.* Dans CMLF2020 : 7e Congrès Mondial de Linguistique Française. Montpellier, France.

Publications dans des actes de colloques nationaux

Federzoni, S. (2020). *Typologie de chaînes de référence à la lumière de corpus annotés diversifiés*. Dans Christophe Benzitoun, Laurine Huber (dir.), 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Communications des apprenti-e-s chercheur-euse-s 2020. Nancy, France.

Communications dans des colloques internationaux

Federzoni, S., Ho-Dac, L.-M. et Fabre C. (2021). *Impact du type de référent sur la composition des chaînes référentielles*. Colloque international Langues et Discours (LED 2021) : "La référence : (co-)construction et exploitation)". Université Grenoble Alpes.

Federzoni, S., Ho-Dac, L.-M. et Fabre, C. (2021). *Mining annotated corpora for coreference chains patterns*. Discourse in Corpus and Experimental Data (DisCorX 2021) : "Bridging the methodological gap". À distance.

Condamines, A., Escoubas Benveniste, MP. et Federzoni, S.(2021). *Étude des marqueurs de méronymie dans un corpus aligné français-italien*. Congrès international de l'Association Française de Linguistique Appliquée, "Outils et nouvelles explorations de la linguistique appliquée (ONELA). Toulouse, France.

Ho-Dac, L.-M., Federzoni, S., Rebeyrolle, J. et Bard, Y. (2021). *Analysis of the reference cohesive ties in students' original draft narratives*. DiscourseNet Conference (DNC4) : "The Impact of Discourse Studies on the Contemporary World". Budapest, Hongrie.

Ho-Dac, L.-M., Federzoni, S., Jacques, M-P., Pallanti, L. et Rebeyrolle, J. (2021). *Chaînes de référence en corpus scolaire et académique : annotation et interprétation*. Colloque international Langues et Discours (LED 2021) : "La référence : (co-)construction et exploitation)". Université Grenoble Alpes.

Ho-Dac, L.-M., Garcia-Debanc, C., Federzoni, S. et Rebeyrolle, J. (2021). *Coreference chains annotation in learners' manuscripts*. Discourse in Corpus and Experimental Data (DisCorX 2021) : "Bridging the methodological gap". À distance.

Ho-Dac, L.-M., Federzoni, S., Bras, M., Rebeyrolle, J. et Garcia-Debanc, C. (2019). *ResolCo un corpus de manuscrits d'élèves et d'étudiants pour l'étude de la cohérence*. 10èmes Journées Internationales de la Linguistique de Corpus. Grenoble, France.

Doquet, C., Federzoni, S., Fleury, S., Ho-Dac, L-M., Mazziotti, S., Moysan, A. et Ponton, C. (2019). *The É :Calm Resource : Transcription, Encoding and Annotation of Handwritten Manuscripts produced by French Pupils and Students* [Poster]. Annotation of non-standard corpora : Prospects and challenges. Bamberg, Allemagne.

Condamines, A., Escoubas-Benveniste, MP. et Federzoni, S. (2018, février). *Usare i corpora allineati per migliorare la costituzione di "reti terminologiche" e l'insegnamento della traduzione (Utiliser des corpus alignés pour améliorer la constitution de réseaux terminologiques et l'enseignement de la traduction)* [Poster]. XVIII Congrès International AItLA. Université de Roma Tre.

Communications dans des colloques nationaux

Federzoni, S. (2019, mai). *Indice ou maillon ? Vers une caractérisation des chaînes de référence à travers l'intuition des locuteurs* [Poster]. Jounrée d'étude "Printemps des Jeunes Chercheurs CLLE". Toulouse.

Séminaires invités

Federzoni, S. (2024). *Typologie de chaînes de référence à la lumière de corpus annotés diversifiés*. Journée d'étude "Les sciences des données et leur utilisation en sciences humaines et sociales". MSH Val de Loire. Orléans, France.

Federzoni, S. (2023). *Études sur les chaînes de référence : approches computationnelles*. Séminaire profésionnel du master en Sciences du Langage, parcours Linguistique outillée et Traitement Automatique des Langues (LouTAL). Université d'Orléans, France.

Federzoni, S. (2020). *Chaînes et maillons dans AnnoDis : référents humains versus référents non humains*. Journée d'étude "Annotation de la coréférence". Toulouse, France.

Ho-Dac, L.-M., Rebeyrolle, J., Garcia-Debanc, C. et Federzoni, S. (2020). *De AnnoDis à É :Calm : d'un guide d'annotation de la co-référence à l'autre*. Journée d'étude "Annotation de la coréférence". Toulouse, France.

Autres activités

2023 **Membre du comité de rédaction de la newsletter du laboratoire CLLE**, Université Toulouse Jean Jaurès

2022-2023 **Membre du collectif CEPEL (Cercle Étudiant Pour l'Étude du Langage)**, Université Toulouse Jean Jaurès

— Aide à l'organisation de séminaires de vulgarisation scientifique, Université Toulouse Jean Jaurès

1-5 juillet 2019 **Membre du comité d'organisation pour la conférence sur le Traitement Automatique des Langues Naturelles (TALN-RECITAL)**, PFIA (Plate-Forme Intelligence Artificielle)

Langues

Italien	*Langue maternelle*
Français	*Niveau C2*
Anglais	*Niveau B1/B2*
Espagnol	*Niveau B2*

Programmation

Pyhton

R ■■■□□

Java ■□□□□

Autres compétences informatiques

Langage de balisage : XML, HTML

Normes : TEI-P5, schémas d'annotation de corpus

Outils d'annotation de corpus : GLOZZ, TXM, Inception, SACR

Outils d'alignement de corpus : Alinéa

Outils d'exploration de corpus : AntConc, TXM

Bureautique : LaTeX, Gimp, MSExcell, Pages, Numbers