

Lydia-Mai Ho-Dac

Maître de Conférence en Sciences du Langage

Université de Toulouse Jean Jaurès
CLLE, UMR 5263
Née en 1976, France
✉ hodac@univ-tlse2.fr



Actuellement

Maître de Conférence à l'Université Toulouse Jean Jaurès, département des Sciences du Langage

Licence Sciences du Langage, Parcours "centré" et Parcours "professorat des écoles"

Master Sciences du Langage, Parcours "Linguistique, Informatique, Technologies du Langage (LITL)"

Membre du laboratoire Cognition, Langues, Langage, Ergonomie (CLLE) UMR 5263

Équipe PLCProcessus Langagiers & Cognitifs

thématique Outils numériques : aspects cognitifs et langagiers

Domaines de recherche

organisation du discours, cohésion textuelle, linguistique de corpus et linguistique outillée, traitement automatique des langues, genres textuels, écrits non standards (communications médiées par les réseaux et écrits d'apprenants)

Formation Universitaire

2000–2007 **Doctorat en Sciences du Langage – TAL**, Université de Toulouse-le Mirail, "La position initiale dans l'organisation du discours : une exploration en corpus" (obtenu le 9-11-2007 et disponible sous [tel \(tel-00176747\)](tel:00176747)). Directrice : M.-P. Péry-Woodley – CLLÉ (équipe ERSS) / Université de Toulouse-Le Mirail. Mention très honorable avec félicitations du jury, *mention très honorable avec félicitations du jury*.

1998–2000 **Master en Sciences du Langage – TAL**, Université de Toulouse-le Mirail, *mention B*.

1995–1998 **Licence Sciences du Langage – FLE**, Université de Nancy 2, *mention AB*.

1993 **Baccalauréat scientifique (section C)**, Lycée J. Poncelet, Saint-Avoid, .

Projets et collaborations

2018–2022 **E-CALM, Écriture scolaire et universitaire : Corpus, Analyses Linguistiques, Modélisations didactiques (E :CALM)**, projet ANR générique, Défi 8 : « Corpus et outils pour la recherche en SHS », Coord. C. Doquet (Clesthia, Paris8).

Structuration, mise à disposition et analyse d'un vaste corpus d'écrits scolaires annotés pour permettre une caractérisation quantitative et outillée de l'évolution des compétences d'écriture au niveau de l'orthographe et de la cohérence.

2021–2023 **Desarrollo y validación de un sistema automatizado de revisión de ítems**, projet FONDEF IDEA 2021, Université du Chili / Université Toulouse Jean Jaurès, coordonné par S. Lions (Université du Chili).

Mise en place d'un système automatique d'évaluation de la qualité de rédaction et de construction des QCM.

2020–2022 **L'apport des méthodes distributionnelles à la sémantique lexicale : analyse morphosémantique et typologie des noms**, Partenariat Hubert Curien / Germaine de Staël, Coord. C. Fabre (Université Toulouse Jean Jaurès) et R. Huyghe (Université de Fribourg).

Exploitation des méthodes distributionnelles automatiques dans la recherche en sémantique lexicale, et plus spécifiquement en sémantique nominale.

2014–2018 **TextLink, Structuring Discourse in Multilingual Europe**, action COST (European Cooperation in Science and Technology), Coord. Liesbeth Degand (UCL Louvain La Neuve, Belgique).

Mise en commun et l'interopérabilité de ressources linguistiques et d'outils d'annotation pour l'étude de la structuration du discours.

2014–2016 **Blog, comment and discuss !**, projet TIAS (Turku Institute for Advanced Studies), collaboration avec Véronika Laippala (Université de Turku, Finlande).

Quantitative study of French internet texts using automatic morpha-syntactic analysis.

- 2014–2018 **RESTAURE, RESsources informatisées et Traitement AUTomatique pour les langues REgionales**, projet ANR, Coord. D. Bernhard (LiLPa, Université de Strasbourg).
Constitution de ressources informatiques et d'outils de traitement automatique pour trois langues régionales de France : alsacien, occitan et picard.
- 2013–2016 **ORFEO, Outils et Recherches sur le Français Ecrit et Oral**, projet ANR : « *Corpus et outils pour la recherche en SHS* », Coord. J.-M. Debaisieux (Lattice).
Constitution d'une plateforme pour l'étude du Français Contemporain, écrit et oral.
- 2013–2015 **Patient's mind, À quoi peuvent bien penser les patients?**, projet interMSH (soumission ANR 2016), Coord. S. Bringay (LIRMM, Montpellier).
Extraction de connaissances et catégorisation des messages dans les forums de santé.
- 2009–2011 **Intermede**, *Projet financé par Institut de Recherche en Santé Publique*, Coord. T. Lang (LISST, U558, Université de Toulouse 2).
Analyse linguistique des interactions qui se produisent entre le patient et le médecin généraliste dans le cadre de consultations médicales pour identifier des inégalités sociales de santé
- 2009–2010 **Journalisme OnLine**, *Projet d'Action de Recherche Concertée – ARC*, Université Catholique de Louvain-la-Neuve et Université Libre de Namur, Collaboration avec L. Degand et A-C Simon (VALIBEL, UCL).
The transformation of the relationship with information in multimedia communication, Analyse semi-automatique des nouveaux « genres » que représentent les nouveaux médias en ligne.
- 2008–2009 **ANNODIS - ANNOtation DIScursive**, *Projet ANR : « Corpus et outils pour la recherche en SHS »*, Coord. M.-P. Péry-Woodley (CLLE-ERSS).
Constitution et mise à disposition d'un corpus français de textes écrits annotés discursivement et d'outils d'aide à l'annotation et à l'exploitation des données annotées.
- 2002-2005 **VIZU**, *Visualisation dynamique des Textes, extraction sélective, affichage spatial multi-échelle et observation des stratégies de lecture*, projet Société de l'information. Partenariat : ERSS (Toulouse), LIMSI (Paris).
- 2002-2005 **GeoSem**, *Traitements sémantiques pour l'Information Géographique : textes, cartes, graphiques*, projet Société de l'information. Partenariat : GREYC (Caen), ERSS (Toulouse), ESO (Caen), MEDIA (Lausanne), 01-05).
- 2001-2005 **ASTD**, *Adverbiaux spatio-temporels et discours*, projet ILF. Partenariat : ERSS (Toulouse), LATTICE (Paris).
- 2001-2004 **Segmenter pour interpréter**, *Étude des marqueurs de segmentation et d'organisation des textes*, projet Cognitique (ERSS :Toulouse). *Projet qui a donné lieu à l'allocation de recherche qui a financé ma thèse.*

Références

[*] Liste complète des publications sous HAL : <https://cv.archives-ouvertes.fr/hodac1m>

Sélection de publications des plus récentes aux plus anciennes

- [1] S. Federzoni, L.-M. Ho-Dac et C. Fabre Coreference Chains Categorization by Sequence Clustering. *2nd Workshop on Computational Approaches to Discourse*, Nov 2021, Punta Cana, Dominican Republic. pp.52-57.
- [2] C. Garcia-Debanc, J. Rebeyrolle et L.-M. Ho-Dac La continuité référentielle dans le corpus RÉVOLCO : méthode d'annotation et premières analyses », *Langue française*, vol. 211, no. 3, 2021, pp. 99-114.
- [3] L.-M. Ho-Dac, S. Fleury et C. Ponton. E :Calm Resource : a Resource for Studying Texts Produced by French Pupils and Students *Language Resources and Evaluation Conference, LREC*, 12, May 2020, Marseille, France. pp.4327-4332.
- [4] L.-M. Ho-Dac, A. Miletic, M. Wauquier et C. Fabre Approches outillées pour l'étude des noms sous-spécifiés ou noms capsules. *Le Français Moderne - Revue de linguistique Française, CILF (conseil international de la langue française)*, 2020.
- [5] L.-M. Ho-Dac et Aleksandra Miletic Analyse contrastive des noms sous-spécifiés à l'oral et à l'écrit à partir d'une extraction automatique. *Langages*, 219, 2020.
- [6] Céline Poudat et L.-M. Ho-Dac Désaccords et conflits dans le Wikipédia francophone. *Travaux linguistiques du CerLiCO*, 29, 2019.
- [7] M.-P. Péry-Woodley, L.-M. Ho-Dac, J. Rebeyrolle, L. Tanguy and C. Fabre. A corpus-driven approach to discourse organisation : from cues to complex markers. *Dialogue & Discourse*, 8, 2017.
- [8] C. Garcia-Debanc, L.-M. Ho-Dac, M. Bras and J. Rebeyrolle. Vers l'annotation discursive de textes d'élèves. *Corpus*, 16, 2017.
- [9] N. Asher, P. Muller, M. Bras, L.-M. Ho-Dac, F. Benamara, S. Afantenos and L. Vieu. ANNODIS and related projects : case studies on the annotation of discourse structure. In Nancy Ide and James Pustejovsky (eds), *Handbook of Linguistic Annotation*, Springer Netherlands, pp.1241-1264, 2017,
- [10] L.-M. Ho-Dac, V. Laippala, C. Poudat and L. Tanguy. French Wikipedia Talk Pages : Profiling and Conflict Detection. In Darja Fišer and Michael Beißwenger (eds), *Investigating Computer-Mediated Communication : Corpus-Based Approaches to Language in the Digital World*, Ljubljana University Press, Faculty of Arts, pp.146-168, 2017.
- [11] L.-M. Ho-Dac and C. Fabre and M.-P. Péry-Woodley and J. Rebeyrolle, and L. Tanguy. On the signalling of multi-level discourse structures. *Discours*, 10, 2012.

- [12] M.-P. Péry-Woodley, S. Afantenos, L.-M. Ho-Dac, and N. Asher. Le corpus annodis, un corpus enrichi d'annotations discursives. *TAL*, 52(3) :71–101, 2011.
- [13] L. Tanguy, C. Fabre, L.-M. Ho-Dac, and J. Rebeyrolle. Caractérisation des échanges entre patients et médecins : approche outillée d'un corpus de consultations médicales. *Corpus, varia*, 10 :137–154, 2011.
- [14] L.-M. Ho-Dac and A. Küppers. La subjectivité à travers les médias : étude comparée de les médias participatifs et de la presse traditionnelle. *Corpus, varia*, 10 :179–200, 2011.
- [15] L.-M. Ho-Dac. An exploratory data-driven analysis for describing discourse organization. In Almela Moisés and Aquilino Sánchez, editors, *A Mosaic of Corpus Linguistics. Selected Approaches*, pages 79–100. Frankfurt/Berlin : Peter Lang, 2010.
- [16] L.-M. Ho-Dac and M.-P. Péry-Woodley. A data-driven study of temporal adverbials as discourse segmentation markers. *Discours*, 4, 2009.

Production et mise à disposition de ressources

- ResolCo <http://redac.univ-tlse2.fr/corpus/resolco.html>, *Corpus de manuscrits d'élèves et d'étudiants produits en réponse à une consigne d'écriture conçue pour mettre en oeuvre de stratégies de résolution de problèmes de cohésion. Le corpus comprend des annotations manuelles concernant le processus d'écriture, les variantes orthographiques observées et certaines structures discursives.*, Ce corpus a été constitué dans le cadre du projet ANR E-CALM (<http://e-calm.huma-num.fr/>).
- E-CALM <https://hdl.handle.net/11403/e-calm/v2>, *Corpus de transcriptions d'écrits d'élèves et d'étudiants (du CP au lycée) encodé selon la TEI-P5 incluant l'annotation des traces d'écritures (ratures, insertions, etc.)*. , Ce corpus a été constitué dans le cadre du projet ANR E-CALM (<http://e-calm.huma-num.fr/>).
- EFG Wikipedia Corpus *Bientôt en ligne*, *Corpus regroupant les articles et pages de discussion des Wikipédia francophone, germanophone et anglophone.*, Collaboration avec C. Poudat (Université Nice Côte d'Azur) et Harald Lungen (IDS, Mannheim).
- WikiDisc <https://hdl.handle.net/11403/wikidisc/v1.1.>, *Corpus de discussions en ligne issues des coulisses de l'encyclopédie Wikipedia francophone.*, Collaboration avec V. Laippala (University of Turku, Finlande).
- AnnoDis <http://redac.univ-tlse2.fr/corpus/annodis.html>, *Corpus de français écrit enrichi d'annotations discursives : relations rhétoriques et structures multi-échelles (chaînes topicales et structures énumératives)*. , Ce corpus a été constitué dans le cadre du projet ANR AnnoDis..
- GeoPo <http://redac.univ-tlse2.fr/corpus/geopo.html>, *Corpus de d'articles expositifs (informatifs et argumentatifs) proposant des réflexions relatives à notre monde d'aujourd'hui (la crise pétrolière, la guerre contre "l'axe du mal", le terrorisme, l'explosion chinoise, la paix au moyen-orient, etc.). Les textes ont été publiés par l'IFRI (www.ifri.org, think tank français indépendant sur les questions internationales) puis annotés syntaxiquement et discursivement au niveau de la position initiale.*

Enseignement

Domaines d'enseignement

- Enseignements introductifs aux sciences du langage, à la démarche scientifique en linguistique, à la linguistique outillée et au Traitement Automatique des Langues (**parcours général**, parcours SDL-PE (Professorat des Écoles), licence MIASHS (Mathématiques et informatique appliquées aux sciences humaines et sociales))
- Enseignements spécialisés en TAL (Master Sciences du langage, **parcours LITL – Linguistique, Informatique, Technologies du langage**)
- Mise en place d'**activité pédagogiques multimédia** (exercices en ligne, construction collaborative de wikis, mise en place de glossaire)

Détails des enseignements

- L1 Méthodologie du travail universitaire
 Les données en SDL (Sciences Du Langage)
 Sociolinguistique
 Origines et Développement du Langage
 Métiers des Sciences du Langage
- L2 Introduction aux spécificités du rapport discours oral / discours écrit dans l'enseignement à l'école primaire
 Collecte de données en SDL
- L3 Analyse de corpus

Linguistique outillée et Traitement automatique du langage

Introduction aux SDL et au TAL pour les étudiants du département de Maths Info

M1 Ressources, Outils et Applications du TAL

M2 Programmation Python adaptée au format XML

Thématiques actuelles de la recherche en TAL

Les documents structurés

M1/M2 Projets tutorés en TAL dont certains ont pu donner lieu à une publication :

- L. Tanguy, L.-M. Ho-Dac, C. Fabre, et al. LITL at SMM4H : an old-school feature-based classifier for identifying adverse effects in Tweets. *5th Social Media Mining for Health Applications Workshop & Shared Task*, Dec 2020, Barcelona, Spain. pp.134–137.
- L.-M. Ho-Dac, L. Tanguy, C. Fabre, et al. LITL at CLEF eHealth2017 : automatic classification of death reports. *CLEF eHealth 2017*, Sep 2017, Dublin, Ireland. (hal-01702705)
- L.-M. Ho-Dac, L. Tanguy, C. Fabre, et al. LITL at CLEF eHealth2016 : recognizing entities in French biomedical documents. *CLEF eHealth 2016*, Sep 2016, Evora, Portugal. (hal-01365928)

M1/M2 Encadrement de projets de recherche ou stage en entreprise :

- 22 mémoires de M1 (Master SDL Parcours "LITL – Linguistique, Informatique, Technologies du langage" et "Linguistique et Cognition")
- 2 mémoires de recherche de M2 (Master SDL Parcours "LITL – Linguistique, Informatique, Technologies du langage")
- 18 rapports de stage de M2 (Master SDL Parcours "LITL – Linguistique, Informatique, Technologies du langage" de l'Université Toulouse Jean Jaurès et "Linguistiques de corpus et outils informatiques" de l'Université de Bordeaux Montaigne)

Voir le détail sur <https://c11e.univ-tlse2.fr/accueil/annuaire/lydia-mai-ho-dac/>

- + Programme d'échange avec l'Université de Belgrade (2017)
- + Programme d'échange avec l'Université de Turku (2018)
- + Programme d'échange avec l'Université de Mannheim (2019)

Encadrement de thèse

2019-2023 Silvia Federzoni (en co-direction avec C. Fabre) *Typologie des chaînes de référence à la lumière de corpus annotés diversifiés*, Financement : CDU.

Responsabilités scientifiques et animation de la recherche

- **Consortium CORLI** (<https://corli.huma-num.fr/>) – **Corpus, Langues et Interactions, TGIR Huma-Num**
Membre du comité de pilotage depuis 2019.
Responsable du groupe de travail sur l'annotation pour le consortium CORLI – Corpus, Langues et Interactions, TGIR Huma-Num, depuis 2019.
- **Revue Discours** (<https://journals.openedition.org/discours/>)
(Co-)Rédactrice en chef de 2018 à 2021 (avec J. Rebeyrolle puis Nicolas Hernandez)
Membre du comité éditorial depuis octobre 2009
Responsable du site web
Coordinatrice éditoriale des numéros 26 (<https://journals.openedition.org/discours/10646>), 16 de la revue Discours (<https://journals.openedition.org/discours/8597>) et 10 (<https://journals.openedition.org/discours/8597>)
- **Annotation de la coréférence** (Toulouse, nov. 2020) : journée d'étude organisée avec J. Rebeyrolle, Claudine Garcia-Debanc et M.-P. Jacques, <https://hal.archives-ouvertes.fr/JECOREF2020>
- **Consortium "TextLink", projet COST (2014–2018)**
Membre depuis 2014
Organisation avec Philippe Muller du colloque international de clôture du projet (Toulouse, 19-21 Mars 2018), <http://textlink.ii.metu.edu.tr/final-action-conference>
L.-M. Ho-Dac et Philippe Muller. *Cross-Linguistic Discourse Annotation : applications and perspectives*. TextLink2018 – Final Action Conference. 2018. (hal-02048987)

- **Multidisciplinary Approaches to Discourse 2010 (MAD 10)** (Moissac, Mars 2010) : Workshop international organisé avec Julie Lemarié, M.-P. Péry-Woodley et Marianne Vergez-Couret
L.-M. Ho-Dac, Julie Lemarié, M.-P. Péry-Woodley et Marianne Vergez-Couret. *Proceedings of the 8th MAD Multidisciplinary Perspectives on Signalling Text Organisation*. MAD – Multidisciplinary Perspectives on Signalling Text Organisation, Mar 2010, Moissac, France. 2010. (hal-01391515)
- Responsable de l'organisation des séminaires du laboratoire CLLE-ERSS de 2012 à 2017
- Animation d'un atelier sur l'annotation de l'écrit dans le cadre de l'école thématique CNRS "Annotation de données langagières" (<http://annotationlinguistique.fr>, Biarritz, sept. 2011) organisé par S. Leroy et T. Poibeau

Responsabilités universitaires

- Responsable de la Licence de Sciences du Langage depuis 2021
- Responsable de la Première année de Licence de Sciences du Langage depuis 2018
- Webmaster du site du département de Sciences du Langage depuis 2012
- Coordinatrice pour les Relations Internationales – mobilités sortantes de 2016 à 2020
- Correspondante RECAP – Responsabilité d'Études et de Coordination Pédagogique pour le département des Sciences du Langage depuis 2020
- Correspondante SCUIO-IP – Service Commun Universitaire d'Information, d'Orientation et d'Insertion Professionnelle pour le département des Sciences du Langage depuis 2013
- Membre du CSQ de 2014 à 2018

Évaluation pour la recherche scientifique

Suivis et comités de thèse

- Jury pour la thèse de Anaëlle Balendent, candidat au doctorat en informatique (Université de Caen Normandie). Titre de la thèse : "De la complexité de l'annotation manuelle : méthodologie, biais et recommandations". Dec. 2022
- Comité de suivi de thèse d'Eleni Metheniti pour l'obtention d'un doctorat en Informatique et SDL (Université Toulouse Jean Jaurès, CLLE et Université Toulouse Paul Sabatier, IRIT). Titre de la thèse : "Construction de modèles sémantiques compositionnels linguistiquement interprétables". Sept. 2021.
- Jury pour la thèse de Ludovic Bocken, candidat au doctorat en informatique cognitive (Université de Québec à Montréal). Titre de la thèse : "Normalisation des articles Wikipédia selon la constance lexicale de leurs introductions et corps de textes". Janv. 2021
- Comité de sélection CLLE-ERSS des candidats au financement CDU, 2017, 2018.
- Jury pour le **Prix de la thèse 2016**, ATALA.

Évaluation de projets scientifiques

- Projet **Emergence(s) 2016**, Mairie de Paris
- Projet ANR CODIM - AAPG2021, phase 1 uniquement

Expertise de publications scientifiques

- Conférence **ICLC-10** 10th International Contrastive Linguistics Conference, 2023, Mannheim, Germany
- Collective volume on "Reference : From Conventions to Pragmatics" (John Benjamins)
- Revue **Discours** numéros D30 (2022); D25 (2019); D21 (2017)
- Revue **Langue Française** numéro thématique "Écrire de l'école à l'université : corpus, traitements, analyses outillées", 211, 2021
- Revue **Lexique** numéro 27, 2020
- Challenge **SMM4H**, Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task, 2020, Barcelona, Spain
- Conférence **JEP-TALN-RECITAL**, 2020, Nancy
- Conférence **EMNLP** Conference on Empirical Methods in Natural Language Processing, 2020, virtual; 2022, Abu Dhabi, UAE
- Conférence **ACL** Annual Meeting of the Association for Computational Linguistics, 2018, Melbourne, Australia

- Conférence **JLC** : Journées de Linguistique de corpus, 2018, 2023, Grenoble
- Conférence **DISROM** : 5th International Conference Discourse Markers in Romance Languages : Boundaries and Interfaces, 2017, Louvain-La-Neuve
- Revue **Corpus** numéro thématique 'Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement', 2018
- Revue **TAL** numéro thématique 'Sémantique distributionnelle', 56(2), 2015
- Revue **Languages in Contrast**, numéro thématique 'Genre- and Register-Related Text and Discourse Features in Contrast', 14(1), 2014
- Workshop **LAW VIII 2014** '8th Linguistic Annotation Workshop', 2014, Dublin, Irlande
- Conférence **ESFLC 2014** (25th European Systemic Functional Linguistics Conference, 2014, Paris
- Revue **Discourse, Context & Media** (G. Myers ed), Elsevier Editor <http://www.journals.elsevier.com/discourse-context-and-media/>
- Workshop **TALaRE 2015** : Traitement Automatique des Langues Régionales de France et d'Europe, 2015, Caen
- Conférence **JETOU** : Les interfaces en Sciences du LangageLe(s), 2017 ; Le(s) discours en Sciences du Langage : Unités et niveaux d'analyse, 2015 ; Variation and Variability in the Language Sciences : analysing, measuring, contextualising, 2013
- Conférence **International Conference on Online Journalism** 2012, Louvain-la-Neuve, Belgique
- Revue **Corpora and Language in Use**, Actes du colloque LPTS 2011 "Across the line of Speech and Writing Variation / Variation(s) sur la structure de l'oral et l'écrit", Presses universitaires de Louvain La Neuve

Expériences professionnelles précédentes

- 2011–2012 **ATER**, *Université de Toulouse-le Mirail, Dpt des Sciences du Langage*, Toulouse.
Poste à temps plein, enseignements en L1, L3 et Master, parcours ECIL – Ergonomie Cognitive et Ingénierie Linguistique : Introduction à l'ingénierie linguistique.
- 2011 **Ingénieure linguiste**, *Atchik-service*, <http://www.atchik-services.com/>, Toulouse.
Mise en place d'un système d'assistance à la modération des forum de discussion par apprentissage automatique supervisé
- 2010–2011 **Post-doctorat – 6 mois**, *CLLE-ERSS*, Toulouse.
projet Intermede : Projet Services de santé de l'Institut de Recherche en Santé Publique : Interactions médecin-patient en médecine générale et inégalités sociales de santé. Analyses interdisciplinaires. Partenariat : U558 INSERM (Toulouse), LISST UMR 5193 (Toulouse), laboratoire de santé publique et d'épidémiologie (Nantes), LERASS (Toulouse), SOI EA 3690 (Toulouse)
- 2009–2010 **Post-doctorat – 12 mois**, *VALIBEL, Université Catholique de Louvain-la-Neuve*, Belgique.
Projet d'Action de Recherche Concertée – ARC : *The transformation of the relationship with information in multimedia communication*, Analyse semi-automatique des nouveaux « genres » que représentent les nouveaux médias en ligne. Partenariat : Université de Louvain-La-Neuve et l'Université de Namur.
- 2008–2009 **Post-doctorat – 12 mois**, *CLLE-ERSS*, Toulouse.
projet ANR , Programme SHS : « Corpus et outils pour la recherche en SHS » : ANNODIS - ANNOtation DIScursive, corpus de référence pour le français et outils d'aide à l'annotation et à l'exploitation. Partenariat : CLLE-ERSS (Toulouse), IRIT (Toulouse) et le GREYC (Caen).
- 2006–2010 **Vacations sporadiques – 6 mois**, *CLLE-ERSS et ATILF*, Toulouse et Nancy.
TELOC : constitution d'une base de données textuelles en Occitan. Partenariat : CNRTL (Nancy), le CROM (Toulouse), CIEL d'ÒC (Toulouse).
- 2006–2007 **Formatrice**, *Société de formation ADONIS*, Toulouse.
Analyse grammaticale (préparation aux concours d'orthophoniste et de professeur des écoles, 2006/2007)
- 2005–2006 **ATER**, *Université de Toulouse-le Mirail, Dpt des Sciences du Langage*, Toulouse.
Poste à mi-temps, enseignements en L1 Sciences du Langage.
- 2000–2004 **Allocataire Monitrice**, *Université de Toulouse-le Mirail, Dpt des Sciences du Langage*, Toulouse, Programme Cognitique, allocation de recherche ministérielle. .
Recherche doctorale et enseignements en L1 Sciences du Langage.
- 2000 **Stage de Master (stagiaire (9mois) + cadre (3mois))**, *EADS*, Toulouse.
Élaboration et évaluation d'un langage contrôlé pour les alarmes en vol (A380).

Langues

Français	Langue maternelle
Anglais	Bon niveau écrit et oral
Allemand	Niveau débutant/moyen
Vietnamien	Niveau débutant/moyen

Compétences informatiques

- **Systèmes d'exploitation** : Linux, Windows
- **Programmation** : Python, Perl, R, Visual Basic, Prolog
- **Langages de balisage** : HTML, XML
- **Normes** : TEI-P5 – méta-données et schémas d'annotation
- **Outils d'annotation de corpus** : GLOZZ, Analec, Inception, Gate
- **Outils d'exploration de corpus** : Antconc, TXM, (No)SketchEngine
- **Bureautique** : MSAccess, MSExcel, LaTeX, Gimp

Formation continue

MSAccess	Gestion de base de données et programmation en Visual Basic , <i>EADS-AirBus</i> , Toulouse, 7x8h, 2000.
HTML	Création de pages Web et Initiation au langage HTML , <i>stage de monitorat</i> , Toulouse, 1x6h, 2003.
XML et XSLT	balisage de documents et transformation (F. Sajous, CLLE-ERSS) , <i>CLLE-ERSS</i> , Toulouse, 2x4h, 2004/2005.
HALshs	Journées sur les archives ouvertes (D. Charnay, CCSD-CNRS et C. Berthaud, service d'ingénierie documentaire et ISH) , <i>Université de Toulouse-Le Mirail</i> , Toulouse, 1x6h, 2008.
TEI-P5	Journée TEI-Council , <i>ENS - SHS</i> , Lyon, 1x8h, 2009.
Anglais	Rédaction de textes scientifiques (A. Pham, UCL) , <i>UCL</i> , Louvain-la-Neuve, Belgique, 12x4h, 2010.
COMPIL	Journée COMPIL sur les Langages de Scripts , <i>Université de Toulouse-Le Mirail</i> , Toulouse, 1x8h, 2010.
Statistiques	Formation CNRS « Statistiques appliquées aux Sciences Humaines » (H. Giraudeau, CLLE-ERSS) , <i>Université de Toulouse-Le Mirail</i> , Toulouse, 9x2h, 2011.

Par ailleurs...

Musicienne, Marionnettiste et Comédienne

1999–2010 **Les Poupées Barbares**, <https://www.lecloudanslaplanche.com/les-poupees-barbares/>

Membre de l'association **Les Vidéophages** : lesvideophages.free.fr/