

Lydia-Mai Ho-Dac

Maître de Conférence en Sciences du Langage

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



Domaines de recherche : linguistique de corpus, TAL, structures discursives, communications médiées par les réseaux (Wikipedia, forum de discussion)

Membre de l'axe CARTEL Corpus, Applications, Ressources pour le Traitement et l'Étude du langage, CLLE-ERSS

Doctorat en linguistique "*La position initiale dans l'organisation du discours : une exploration en corpus*" (obtenu le 9-11-2007 et disponible sous [tel](#)). Directrice : Marie-Paule Péry-Woodley – CLLE-ERSS / Université de Toulouse-Le Mirail. Mention très honorable avec félicitations du jury

Projets et collaborations

- 2018–2021 **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* », coordonné par C. Doquet (Clesthia).
Structuration et mise à disposition d'un vaste corpus d'écrits d'élèves et d'étudiants permettant des analyses quantitatives et des traitements automatiques ; caractérisation de ces écrits en termes d'acquisition des compétences au niveau de l'orthographe et de la cohérence ; étude des modalités d'écriture dans les avant-textes (plans, notes, brouillons) et les textes.
- 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, coordonné par 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* », coordonné par 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*), coordonné par S. Bringay, LIRMM, Montpellier.
Extraction de connaissances et catégorisation des messages dans les forums de santé.
- 2010–2011 **Intermede**, Collaboration avec C. fabre, J. Rebeyrolle, L. Tanguy 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.
- 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* », coordonné par 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.

Références

- [1] Marie-Paule Péry-Woodley, Lydia-Mai Ho-Dac, Josette Rebeyrolle, Ludovic Tanguy and Cécile Fabre. A corpus-driven approach to discourse organisation : from cues to complex markers. *Dialogue & Discourse*, 8, 2017,
- [2] Claudine Garcia-Debanc, Lydia-Mai Ho-Dac, Myriam Bras and Josette Rebeyrolle. Vers l'annotation discursive de textes

d'élèves. *Corpus*, 16, 2017.

- [3] Nicholas Asher, Philippe Muller, Myriam Bras, Lydia-Mai Ho-Dac, Farah Benamara, Stergos Afantenos and Laure 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,
- [4] Lydia-Mai Ho-Dac, Veronika Laippala, Céline Poudat and Ludovic 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.
- [5] Lydia-Mai Ho-Dac and Cécile Fabre and Marie-Paule Péry-Woodley and Josette Rebeyrolle, and Ludovic Tanguy. On the signalling of multi-level discourse structures. *Discours*, 10, 2012.
- [6] Marie-Paule Péry-Woodley, Stergos Afantenos, Lydia-Mai Ho-Dac, and Nicholas Asher. Le corpus annodis, un corpus enrichi d'annotations discursives. *TAL*, 52(3) :71–101, 2011.
- [7] Ludovic Tanguy, Cécile Fabre, Lydia-Mai Ho-Dac, and Josette Rebeyrolle. Caractérisation des échanges entre patients et médecins : approche outillée d'un corpus de consultations médicales. *Corpus, varia*, 10 :137–154, 2011.
- [8] Lydia-Mai Ho-Dac and Anne Küppers. La subjectivité à travers les médias : étude comparée de les médias participatifs et de la presse traditionnelle. *Corpus, varia*, 10 :179–200, 2011.
- [9] Lydia-Mai 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.
- [10] Lydia-Mai Ho-Dac and Marie-Paule Péry-Woodley. A data-driven study of temporal adverbials as discourse segmentation markers. *Discours*, 4, 2009.

Liste complète des publications sous HAL : <https://cv.archives-ouvertes.fr/hodaclm>

Enseignement

Domaines d'enseignement

- Enseignements introductifs au sciences du langage, linguistiques de corpus et Traitement Automatique des Langues (**parcours général**)
- 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
 - Origines et Développement du Langage
 - Outil pour l'étudiant
 - Métiers des Sciences du Langage
- L2 Option pour la préparation au métier de Professeur des écoles
- L3 Analyse de corpus
 - Linguistique et Traitement automatique du langage
- M1 Outils et Applications du TAL
 - Ressources linguistiques pour le TAL
 - Programmation Python adaptée au format XML
 - Collecte et Bases de données
- M1/M2 Projets tutorés en en TAL
 - + Programme d'échange avec l'Univeritsé de Belgrade
 - Introduction au TAL, étiquetage Morpho-syntaxique et analyse syntaxique

Encadrement de mémoires et rapports de stage - Master professionnel

- 2017 Laura Rivière *Identification du genre des auteurs et destinataires dans les profils d'un site de rencontre*, mémoire et rapport de stage de Master 2 LITL, Université de Toulouse Jean-Jaurès. Partenaire : E. Bantyukova, ioSquare
- 2016 Marine Raoult *Étude préliminaire pour l'intégration de modules de synthèse vocale et de reconnaissance vocale dans les futurs cockpits airbus*, mémoire et rapport de stage de Master 2 ECIL, Université de Toulouse Jean-Jaurès. Partenaire : E. Cannesson, AIRBUS

Aglaée Philipi *Influence de l'émotion dans la description d'un lieu*, mémoire et rapport de stage de Master 2 ECIL, Université de Toulouse Jean-Jaurès. Partenaire : C. Domingues, COGIT, IGN et A. Lacheret-Dujour, MODYCO

2015 Valentine Delbar *Outil d'évaluation positive de l'orthographe*, mémoire et rapport de stage de Master 2, Université de Toulouse-Jean-Jaurès. Partenaire : S. Muller, Synapse Développement

2014 Émilie Merdy *Outils de syllabations pour l'aide à la lecture chez de jeunes dyslexiques*, mémoire et rapport de stage de Master 2, Université de Toulouse-Jean-Jaurès. Partenaire : Joël Bachelier et Gilles Vessière, Médialexie

2013 Sarah Growas *Conception d'un environnement informatique pour l'apprentissage humain (EIAH)*, mémoire et rapport de stage de Master 2 co-encadré avec C. Navarro, Université de Toulouse-le Mirail. Partenaire : Lucas Pruet, Alliance Française de Paris

Céline Barès *Amélioration du système de Foire Aux Questions pour le projet client Butagaz*, mémoire et rapport de stage de Master 2 co-encadré avec L. Tanguy, Université de Toulouse-le Mirail. Partenaire : Manon Quintana, Inbenta

Encadrement de mémoires - Master recherche

2017 Anouk Birski *Catégorisation automatique d'actes de communication. Les fils de discussion des pages de discussion Wikipédia*, mémoire de Master 1 co-encadré avec C. Fabre, Université de Toulouse-Jean-Jaurès

Elisa Piccinini *Étude contrastive du vocabulaire médical : Forum en ligne et vulgarisation médicale*, mémoire de Master 1 co-encadré avec C. Fabre, Université de Toulouse-Jean-Jaurès

2016 Heu Mby Aurore *Étude d'un corpus de forme wikipédia pour l'analyse et la détection des accord/désaccords*, mémoire de Master 1 co-encadré avec C. Fabre, Université de Toulouse-Jean-Jaurès

2015 Marine Raoult *Pour un repérage automatique des structures énumératives à l'aide des noms sous-spécifiés*, mémoire de Master 1 co-encadré avec J. Rebeyrolle, Université de Toulouse-Jean-Jaurès

Marie Voisin *La gestion des posts sensibles dans les forums de santé : Expression et détection du risque dans les messages des internautes*, mémoire de Master 1 co-encadré avec N. Souf (IRIT), Université de Toulouse-Jean-Jaurès

2014 Magali Capeyron *Analyse des citations dans les forums de santé*, mémoire de Master 2 (Master ReLai) co-encadré avec N. Souf et M. Mojahid, Université de Bordeaux

Estelle Liaubet *Étude comparative des marqueurs d'attitude dans les discussions wikipedia*, mémoire de Master 1 co-encadré avec C. Fabre, Université de Toulouse-Jean-Jaurès

2013 João Pedro Campello Rodrigues Pereira de Figueiredo *Classification automatique de séquences textuelles de rapports d'accidents aériens : une approche rhétorique*, mémoire de Master 1 co-encadré avec N. Tulechki, Université de Toulouse-le Mirail.

Émilie Merdy *Traitement automatique des rappels de souvenirs épisodiques. Recherches de phénomènes langagiers particuliers chez des sujets sains et épileptiques*, mémoire de Master 1 co-encadré avec L. Tanguy, Université de Toulouse-le Mirail.

Responsabilités scientifiques et animation de la recherche

— Membre du consortium CorLi "Corpus, Langues et Interactions", ILF, TGIR Huma-Num

— Responsable de l'organisation des séminaires du laboratoire CLLE-ERSS de 2012 à 2017

— Membre du consortium "TextLink", projet COST

— Membre du comité éditorial de la revue Discours depuis octobre 2009 (discours.revues.org) et co-responsable du site web avec Stéphane Ferrari

— Coordinatrice éditoriale du numéro 16 de la revue Discours ([discours.revues.org](http://discours.revues.org/8597) (<http://discours.revues.org/8597>))

— Coordinatrice scientifique du numéro thématique 10 de la revue Discours ([discours.revues.org](http://discours.revues.org/8597) (<http://discours.revues.org/8597>))

Responsabilités universitaires

— Co-webmaster pour la page internet du département de Sciences du Langage

— Correspondante pour les Relations Internationales – mobilités sortantes depuis 2016

— Correspondante SCUIO-IP – Service Commun Universitaire d'Information, d'Orientation et d'Insertion Professionnelle pour le département des Sciences du Langage depuis 2013

Expertises de projets et publications scientifiques

- Conférence **DISROM** : 5th International Conference Discourse Markers in Romance Languages : Boundaries and Interfaces, 2017, Louvain-La-Neuve
- Jury pour le **Prix de la thèse 2016**, ATALA
- Projet **Emergence(s) 2016**, Mairie de Paris
- Revue **Corpus** numéro thématique 'Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement', à paraître
- 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-laNeuve, Belgique