

Marine WAUQUIER – CV détaillé

Date et lieu de naissance 23/09/1990 à Charenton-le-Pont (94)
Nationalité Française

Email marine.wauquier@univ-tlse2.fr
Adresse professionnelle Bureau B516, Maison de la Recherche
5 Allées Antonio Machado
31058 Toulouse Cedex 9
France

Page personnelle <https://bit.ly/3hHGGVd>

Formation

Formation universitaire

2017- Doctorat en Sciences du langage

Financé par un Contrat Doctoral Unique de l'Université Toulouse Jean Jaurès
« Confrontation des procédés dérivationnelles et des catégories sémantiques dans les modèles distributionnels »
Sous la direction de Nabil Hathout (CLLE)
CLLE (UMR 5263), Université Toulouse Jean Jaurès & CNRS

2015-2017 Master mention Sciences du langage, parcours Linguistique, Informatique et Technologies du langage

Université Toulouse Jean Jaurès, Mention Très bien

- **Mémoire de M2** – « Explorations méthodologiques pour la caractérisation distributionnelle de dérivés morphologiques », encadré par Cécile Fabre et Nabil Hathout
- **Mémoire de M1** – « Indices distributionnels pour la comparaison sémantique de dérivés morphologiques », encadré par Cécile Fabre et Nabil Hathout
- **Projets tutorés collectifs :**
 - **M2** : Valorisation et transformation automatique d'une ontologie en wiki d'entreprise dans un domaine de spécialité
 - *En collaboration avec Coup de Puce Expansion*
 - **M1** : Participation au challenge scientifique **CLEF eHealth 2016**, visant l'extraction et la catégorisation automatique d'entités nommées médicales du français
 - *Meilleur score de précision pour le corpus EMEA*

2012-2015 Double licence

Université Lille 3, Mention Bien

- Licence mention **Langues, Littératures et Civilisations Etrangères et Régionales**, parcours **Anglais**
- Licence mention **Lettres**, parcours **Lettres Modernes**

Formation continue

2017-2018 Formations aux pratiques pédagogiques pour l'université (50h)

Dans le cadre des 100h de formations dispensées par l'Ecole Doctorale CLESCO

- Boîtes à outils pédagogiques – 14h
- Pourquoi enseigner avec le numérique ? Réfléchir à son enseignement différemment. – 6h
- Comment intégrer le numérique dans son enseignement ? Exemples et mises en application. – 6h
- Découvrir ce qu'est apprendre pour enseigner plus efficacement – 6h
- Maîtriser le corps et l'espace – 6h
- Pratiques théâtrales pour la didactique – 6h
- Droit d'auteur : Enseignement, Recherche – 6h

- 2018** **30th European Summer School in Logic, Language and Information (ESLLI 2018)**
Université St. Kliment Ohridski de Sofia, Bulgarie
 Bourse d'Aide à la Mobilité Internationale des Doctorants (AMID) octroyée par la Commission Recherche de l'Université Toulouse Jean Jaurès
- Word Vector Space Specialisation – Vulić, I. (7,5h)
 - Advanced Regression Methods for Linguistics – Wieling, M. (7,5h)
 - The Lexicon : an Interdisciplinary Introduction – Jezek, E. (7,5h)
 - NLP in the Era of Big Data, Deep Learning, and Post Truth – Nakov, P. (7,5h)
 - Annotation in Digital Humanities (annDH): How Can Linguistics/Computational Linguistics Help with Annotation in DH – Kübler, S., & Zinsmeister, H. (7,5h)
 - Probabilistic Modeling and Bayesian Data Analysis in Experimental Semantics and Pragmatics – Franke, M., & Tessler, M. H. (7,5h)
 - Bridging Formal and Conceptual Semantics – Balogh, K., & Petersen, W. (7,5h)
 - Computational Models of Events – Pustejovsky, J. (7,5h)
 - Treebanks, Linguistic Theories and Applications – Osenova, P., & Simov, K. (7,5h)

- 2017** **29th European Summer School in Logic, Language, and Information (ESLLI 2017)**
Université Toulouse Capitole, France
- Introduction to Categorical Compositional Distributional Semantics – Kartsaklis, D., & Lewis, M. (7,5h)
 - Computational lexical semantics – White, A.S., & Rawlins, K. (7,5h)
 - Embeddings and Deep Learning – Schütze, H. (7,5h)
 - Explicit world-knowledge and distributional semantic representations – Sayeed, A., & Zarcone, A. (7,5h)
 - Introduction to Computational Morphology – Sorokin, A. (7,5h)
 - Visualization for Linguistic Research: Methodology Fundamentals – Scrivner, O. (7,5h)

Stage

- 2016** **Stage dans le laboratoire CLLE (2 mois)**
 Participation au projet ANR RESSources informatisées et Traitement AUTomatique pour les langues Regionales (RESTAURE)
- Constitution et enrichissement de ressources pour le traitement automatique de l'occitan
 - Annotation automatique des toponymes en occitan

Enseignement

- 2019-2020** **Avenant d'enseignement au contrat CDU**
- **64h** équivalent TD au Département des Sciences du langage (SDL), UFR Langues, Littératures et Civilisations Etrangères (LLCE), Université Toulouse Jean Jaurès

Période	Code UE	Intitulé du cours	Niveau	Nature	Effectif	Heures
2019-2020 Semestre 2	SL00401V	Fondements de la morphologie	Licence 2	CM & TD	40	24
	SL00205V	Métiers des Sciences du langage	Licence 1	1 séance	100	2
2019-2020 Semestre 1	SLT0906V	Thématiques actuelles de la recherche en TAL	Master 2	TD	8	6
	SL00504V	Analyse de corpus	Licence 3	CM & TD	30	24

- 2018-2019** **Avenant d'enseignement au contrat CDU**
- **64h** équivalent TD au Département des Sciences du langage (SDL), UFR Langues, Littératures et Civilisations Etrangères (LLCE), Université Toulouse Jean Jaurès

Période	Code UE	Intitulé du cours	Niveau	Nature	Effectif	Heures
2018-2019 Semestre 2	SL00202V	Histoire de la linguistique	Licence 1	CM & TD	50	24
	SL00205V	Métiers des Sciences du langage	Licence 1	1 séance	100	2
2018-2019 Semestre 1	SL00304V	Lexicographie	Licence 2	CM & TD	35	24
	SLT0906V	Thématiques actuelles de la recherche en TAL	Master 2	1 séance	8	2

Encadrement

Co-encadrement de 3 mémoires

2019-2020 Mickaëla Vandewaetere. Mémoire de Master 1 LITL du Département des Sciences du langage de l'Université Toulouse Jean Jaurès. « Sémantique des noms déverbaux d'action suffixés en *-age*, *-ion* et *-ment* : étude du degré de technicité ».

Co-encadrement avec Cécile Fabre.

2018-2019 Mathilde Esperce. Mémoire de Master 1 LITL du Département des Sciences du langage de l'Université Toulouse Jean Jaurès. « Étude des différences sémantiques entre dérivés morphologiques : le cas de la féminisation des noms d'agent suffixés en *-eur* ».

Co-encadrement avec Cécile Fabre.

2017-2018 Frédéric Saudemont. Mémoire de Master 1 LITL du Département des Sciences du langage de l'Université Toulouse Jean Jaurès. « Étude de la stabilité des relations sémantiques lexicales basées sur les modèles sémantiques distributionnels ».

Co-encadrement avec Nabil Hathout.

Recherche

Thèse

Ma thèse porte sur l'utilisation de la sémantique distributionnelle dans la recherche fondamentale en linguistique. Dans une optique méthodologique, j'étudie plus spécifiquement à l'aide de l'outil Word2Vec des noms construits du français : les noms d'agent en *-eur*, *-euse* et *-rice* et les noms d'action en *-age*, *-ion* et *-ment*. Je montre que la sémantique distributionnelle permet de valider, d'explorer et de faire émerger à grande échelle des hypothèses linguistiques à partir de larges corpus. J'explore par exemple la similarité sémantique des noms d'agent et d'action déverbaux vis à vis de leur base par le biais de leurs scores de proximité distributionnelle respectifs. En comparant ces scores pour plus de 2500 triplets de noms d'agent, verbes et noms d'agent, je confirme sur un plan empirique que les verbes tendent généralement à être plus proche de leur nom d'action dérivé que de leur nom d'agent dérivé. Je développe par ailleurs une méthodologie d'analyse des catégories sémantiques formées par les noms d'agent en *-eur*, *-euse* et *-rice* d'une part, et les noms d'action en *-age*, *-ion* et *-ment* d'autre part. Je me base pour cela sur la compositionnalité des représentations vectorielles construites par Word2Vec, et propose une représentation moyenne unique. La comparaison des représentations obtenues pour chaque catégorie me permet de montrer que, sur le plan distributionnel :

- les noms d'agent masculins et féminins se distinguent relativement au genre ;
- les noms en *-euse* et *-rice* se distinguent relativement à des connotations dépréciatives ;
- les noms en *-age*, *-ion* et *-ment* se distinguent relativement au degré de technicité des actions qu'ils dénotent.

Sur la base de ces derniers résultats, j'ai mené une étude de la technicité des noms d'action, notion que j'ai définie puis évaluée à l'aide de critères linguistiques calculés à partir de ressources lexicales. La mise en place de dispositifs expérimentaux à la complexité croissante permet de définir des conditions d'utilisation de l'outil distributionnel, mais aussi d'explorer les possibilités théoriques par la combinaison d'hypothèses et de données contrôlées.

Participation à un projet financé

2020-2022 Programme Germaine de Staël (financement dans le cadre du Partenariat Hubert Curien franco-suisse)

Partenaires CLLE & Université de Fribourg ; durée de 2 ans, montant de 2400 €

« L'apport des méthodes distributionnelles à la sémantique lexicale : analyse morphosémantique et typologie des noms »

- Porté par Richard Huyghe (U. Fribourg) et Cécile Fabre (U. Toulouse Jean Jaurès), ce projet est le prolongement du travail en cours sur les noms d'agent que je mène conjointement avec R. Huyghe. Cette collaboration entamée en 2018 vise l'application des outils de calcul et d'analyse de la sémantique des noms dérivés développés dans ma thèse et basés sur les représentations vectorielles fournies par l'outil Word2Vec à des hypothèses concernant plus largement l'ensemble des noms d'agent. L'objectif de ce projet est de développer davantage la méthodologie de recherche employée, mais aussi à étendre son champ d'application à d'autres catégories nominales, telles que les noms d'événement ou les noms sous-spécifiés. D'une manière générale, le travail projeté bénéficiera à la fois aux études en sémantique lexicale et permettra d'avancer dans la compréhension des outils distributionnels et des moyens de les intégrer dans la recherche en linguistique.

Publications

Articles de revues avec comité de lecture

- Huyghe, R., & **Wauquier, M.** (à paraître). What's in an agent ? A distributional semantics approach to agents nouns in French. *Morphology*.
- Huyghe, R., & **Wauquier, M.** (accepté). Une étude distributionnelle des noms d'agent en *-ant, -eur, -ien, -ier* et *-iste Verbum*.
- Ho-Dac, L.M., Miletić, A., **Wauquier, M.**, & Fabre, C. (2020). Approches outillées pour l'étude des noms sous-spécifiés ou noms capsules. *Le Français Moderne*.
- **Wauquier, M.**, Hathout, N., & Fabre, C. (soumis). Semantic Discrimination of Nominalizations based on Distributional and Statistical Clues. *The Journal of Word Formation/Zeitschrift für Wortbildung*.
- Huyghe, R., & **Wauquier, M.** (soumis). Distributional semantics insights on agentive suffix rivalry in French. *Word Structure*.

Chapitre d'ouvrage

- **Wauquier, M.**, Fabre, C., & Hathout, N. (à paraître). Différenciation des noms d'action dérivés : le facteur de technicité étudié en corpus. In C. Frérot & M. Pecman (eds.), *Des corpus numériques à la modélisation linguistique en langues de spécialité*. Grenoble : UGA Éditions

Articles dans des actes de conférences avec comité de lecture

- **Wauquier, M.**, Hathout, N., & Fabre, C. (2020). Contributions of distributional semantics to the semantic study of French morphologically derived agent nouns. In J. Audring, N. Koutsoukos & C. Manouilidou (Eds), *Rules, patterns, schemas and analogy, MMM12 Online Proceedings*, (12), 111-121.
- **Wauquier, M.**, Fabre, C., & Hathout, N. (2018). Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. In F. Neveu, B. Harmegnies, L. Hriba et S. Prévost (eds.), *Actes du 6e Congrès Mondial de Linguistique Française (CMLF)*, SHS Web Conferences, 46. DOI : 10.1051/shsconf/20184608006.
- **Wauquier, M.** (2018). Analyse des noms agentifs dans les espaces vectoriels distributionnels. In *Actes de la conférence Traitement Automatique de la Langue Naturelle (TALN) 2018*, 27-39.
- Ho-Dac, L. M., Tanguy, L., Grauby, C., Mby, A. H., Malosse, J., Rivière, L., Veltz-Mauclair, A., & **Wauquier, M.** (2016). LITL at CLEF eHealth2016 : recognizing entities in French biomedical documents. In *CLEF eHealth 2016*.

Communications orales

- **Wauquier, M.**, Hathout, N., & Fabre, C. (2019). Semantic discrimination of nominalizations based on distributional and statistical clues. *The 8th Workshop on Nominalizations (JENom 8)*. Stuttgart, Allemagne.
- Huyghe, R., & **Wauquier, M.** (2019). Comment reconnaître un nom d'agent ? Le recours à la sémantique distributionnelle pour la classification nominale. *Le nom, entre syntaxe et sémantique*. Paris, France.
- Huyghe, R., & **Wauquier, M.** (2019). Agent nouns and affix selection - Towards a morphosemantic typology of French agent nouns based on distributional semantics. *Morphological Rivalry and Polysemy Workshop*. Paris, France.

Communications invitées

- **Wauquier, M.** (2019). Approches outillées pour l'étude des noms capsules. *Séminaire de master*. Université de Fribourg, Suisse.
- Huyghe, R., & **Wauquier, M.** (2019). Du corpus au lexique. L'identification des noms d'agent au moyen de la sémantique distributionnelle. *Séminaire Linguistique de corpus*. Université de Neufchatel, Suisse.
- Huyghe, R., & **Wauquier, M.** (2020). Des noms d'agent en français : une approche distributionnelle. *Séminaire d'équipe LiLPa*. Université de Strasbourg, France.

Administration et animation scientifique

Responsabilités collectives

2018-2020 Représentante élue des doctorants CLLE-ERSS

Représentation des doctorants de l'équipe ERSS (35 sur le site toulousain, 10 sur le site bordelais) au sein des instances d'équipe

- Participation aux bureaux de l'équipe ERSS
- Interface avec la direction : répartition des bureaux, renouvellement de matériel, gestion de dossiers courants
- Organisation des journées de rentrée des doctorants CLLE (éditions 2018 et 2019) : communication, gestion du budget, planification, présentation du laboratoire et de son fonctionnement

2017-2020 Membre du collectif Cercle Étudiant pour l'Étude du Langage

Organisation de séminaires de vulgarisation à destination des étudiants de licence et master SDL à l'Université Toulouse Jean Jaurès

- Communication et diffusion
- Logistique – organisation des salles
- Choix des intervenants

Organisation de manifestations scientifiques

Journée d'étude « Printemps des Jeunes Chercheurs CLLE 2020 »

Université Toulouse Jean Jaurès, Toulouse

Membre du comité d'organisation

- Communication : appel à communication
- Logistique : gestion du budget, des salles, contact avec les intervenants, planification

TALN-RECITAL 2019

Université Toulouse Capitole, Toulouse

Bénévole dans le cadre de l'organisation de la conférence TALN-RECITAL hébergée par la Plate-Forme Intelligence Artificielle (PFIA) 2019

- Accueil des participants
- Logistique
- Encadrement technique

Journée d'étude « Printemps des Jeunes Chercheurs CLLE 2019 »

Université Toulouse Jean Jaurès, Toulouse

Membre du comité d'organisation

- Communication : appel à communication, diffusion
- Logistique : gestion du budget, planification

Cross-Linguistic Discourse Annotation : applications and perspectives 2018

IRIT, Toulouse

Bénévole dans le cadre de l'organisation de la conférence internationale clôturant l'action COST TextLink

- Accueil des participants

29th European Summer School in Logic, Language, and Information

Université Toulouse Capitole, Toulouse

Bénévole dans le cadre de l'organisation de l'édition 2017 de l'école d'été ESSLLI

- Accueil des participants
- Logistique
- Encadrement technique

Autres travaux de recherche

2018- Recherches à l'interface entre linguistique et games studies

Travaux ayant donné lieu à des communications

Communications orales

- **Wauquier, M.** (2018). Harry Potter et la reconquête de la qualité vidéoludique – Un voyage entre livre et jeu vidéo. *Littératures du jeu vidéo*. Paris, France.
- **Wauquier, M.** (2019). Existe-t-il un genre de la quête dans Skyrim ? Essai de caractérisation linguistique du contenu textuel de The Elder Scrolls. *Les langages du jeu vidéo : codes, discours et images en jeu*. Lausanne, Suisse.

Chapitre d'ouvrage

- **Wauquier, M.**(2019). Harry Potter et la reconquête du caractère vidéoludique – un voyage entre livres et jeux vidéo. In *Littératures du jeu vidéo, Actes du colloque, ENS Ulm, 15 et 16 juin 2018*. Paris : La Taupe Médite. 153-171.

Langues

Français	Langue maternelle
Anglais	Courant
Espagnol	Scolaire
Occitan, latin, ancien français	Notions

Compétences informatiques

Programmation	Python, Perl, Java
Encodage	HTML, XML, OWL, JSON
Outils de TAL	Word2Vec, Talismane, TreeTagger, Praat
Outils de linguistique outillée	AntConc, TXM, BRAT, Protégé, VUE
Divers	R, LaTeX, SQL, Access, MediaWiki, Processing