

OLGA SEMINCK

olga.seminck@cncrs.fr

Table des matières

1 Curriculum Vitæ	2
1.1 Fonctions d'enseignement et de recherche	2
1.2 Diplômes universitaires	2
1.3 Mobilité Internationale	3
1.4 Formation	3
1.5 Bourses	3
1.6 Prix	3
1.7 Expérience industrielle	4
1.8 Stages	4
1.9 Compétences Informatiques	4
1.10 Langues	4
2 Publications	4
2.1 Articles dans des revues à comité de lecture	5
2.2 Chapitres d'Ouvrages	5
2.3 Articles soumis dans des actes de conférences à comité de lecture	5
2.4 Articles publiés dans des actes de conférences à comité de lecture	5
2.5 Communications sans actes à comité de sélection présentées à des conférences	6
2.6 Ressources disponibles en ligne	7
2.7 Vulgarisation scientifique	7
2.8 Communications présentées dans des journées d'étude sans comité de sélection	7
2.9 Conférences dans des séminaires	8

CURRICULUM VITAE

FONCTIONS D'ENSEIGNEMENT ET DE RECHERCHE

Situation actuelle :

- Ingénieure de Recherche, Centre National de la Recherche Scientifique (CNRS),
Laboratoire Langues, Textes, Traitements informatiques, Cognition (Lattice)
Production, Traitement et Analyse des Données 01/12/2021 - aujourd'hui
- Post-doc, Centre National de la Recherche Scientifique (CNRS),
Laboratoire Langues, Textes, Traitements informatiques, Cognition (Lattice)
Humanités numériques, grandes collections littéraires 01/09/2020 - 30/11/2021
- Congé Maternité 19/02/2020 - 03/06/2020
- Post-doc, Université catholique de Louvain, Belgique 01/09/2019 - 31/08/2020
Institute of Neuroscience, Centre de traitement automatique du langage (Cental),
Media innovation and intelligibility Lab (MiiL), post-doc sur le projet Mind-it
Détection automatique de la maladie d'Alzheimer dans des messages électroniques
- Attachée Temporaire de l'Enseignement et de Recherche, Sorbonne Université 01/09/2018 - 31/08/2019
UFR de Sociologie et informatique pour les sciences humaines, intervenant dans le Master
Langue et Informatique et la Licence *Langue française et techniques informatiques*
- Doctorante contractuelle avec monitorat, Université Paris Diderot, UFR de Linguistique, 01/09/2015 - 31/08/2018
Laboratoire de Linguistique Formelle & Laboratoire d'Informatique de Paris Nord

DIPLÔMES UNIVERSITAIRES

Doctorat

- Thèse en Linguistique Informatique, Université Sorbonne Paris Cité, Université Paris Diderot 23/11/2018
Cognitive Computational Models of Pronoun Resolution.
Rapporteur & Président : Laurent Prévot, Professeur en Sciences du Langage, Université d'Aix-Marseille
Rapporteur : Aline Villavicencio, Lecturer en Informatique, University of Essex
Examineur : Saveria Colonna, Maître de Conférences HDR en Sciences du Langage, Université Paris 8
Examineur : Vera Demberg, Professeur en Informatique, Universität des Saarlandes
Directeur : Pascal Amsili, Maître de Conférences HDR en Sciences du Langage, Université Paris Diderot
Co-directeur : Adeline Nazarenko, Professeur en Informatique, Université Paris 13

Master & Licences

- Master en Linguistique Informatique, Mention *Très Bien*, Université Paris Diderot 26/06/2015
Mémoire Master 2 : *Modèle distributionnel de l'acquisition des noms et des verbes.*
Laboratoire de Linguistique Formelle & Laboratoire de Sciences Cognitives et Psycholinguistique
Encadrement : P. Amsili & A. Christophe
- Licence en Linguistique Informatique, Mention *Très Bien*, Université Paris Diderot 01/07/2013
Mémoire de stage M1 : *Prototype d'un classifieur de tweets à buts communicatifs.*
Laboratoire Modèles, Dynamiques, Corpus (MoDyCo), Université Paris-X
Encadrement : J.L. Minel

· Licence en linguistique générale, deuxième matière langue et culture française,
Cum Laude et Honours Degree, Vrije Universiteit Amsterdam
Mémoire de Licence : *Het Aspect van de Ok-talen (L'aspect dans les langues Ok)*.
Faculté des sciences humaines
Encadrement : L. de Vries

31/08/2012

MOBILITÉ INTERNATIONALE

Mobilité internationale de 3 mois à l'Université de Saarland (Allemagne) dans le groupe *01/03/2018 - 01/06/2018*
Computer Science and Computational Linguistics du Pr. V. Demberg

FORMATION

Statistiques Bayésiennes, CNRS *2022*
Différentes formations de DoRANum (Données de la recherche : Apprentissage Numérique), CNRS *2022*
Présentation de la plateforme DoRANum, Partager et valoriser ses recherches via les réseaux sociaux, L'Outil DMP OPIDoR : modèle de plan de gestion de données structuré, Plan de gestion de données : bonnes pratiques de rédaction et outils
Formation Introduction au Deep Learning (FIDDLE), CNRS *2021 - 2022*
Formation en électroencéphalographie (EEG), *2019*
Laboratoire d'Excellence Empirical Foundations of Linguistics, Paris
Programme de formation doctorale de 300 heures, École doctorale Frontières du Vivant, Paris *2015 - 2018*
Creating Interdisciplinary Research Projects (2015), Engaging Scientific Presentations (2016), Critical Analysis of Research Articles (2016), Scientific Writing (2017), Transforming and Leveraging your Scientific Skills (2017)
Workshop modélisation bayésienne, Laboratoire d'Excellence Empirical Foundations of Linguistics, Paris *2017*
École d'automne grammaire expérimentale, Labex Empirical Foundations of Linguistics, Cargèse *2017*
Lisbon Machine Learning School (LXMLS), Lisbonne *2016*
European Summer School in Logic, Language and Information (ESSLLI), Barcelone *2015*

BOURSES

Bourse Post-doctorale, PRAIRIE (Paris Artificial Intelligence Research Institute) *2020 - 2021*
Bourse Doctorale *Double Culture* (financement de thèse), Université Sorbonne Paris Cité *2015 - 2018*
Bourse Post-doctorale, Université catholique de Louvain *2019 - 2020*
Bourse Doctorale *Double Culture* (financement de thèse), Université Sorbonne Paris Cité *2015 - 2018*
Bourse de Mobilité de 3 mois, Labex Empirical Foundations of Linguistics *2018*
Voir la section 'Mobilité Internationale'.
Bourse ATALA (Association pour le Traitement Automatique des Langues) pour assister *2017*
à la Conférence du Traitement Automatique des Langues Naturelles

PRIX

Meilleur article de la 18ième Rencontre des Étudiants Chercheurs en Informatique *2016*
pour le Traitement Automatique des Langues (RECITAL)
Junior Prof, Prix pour le meilleur mémoire de lycéens dans le domaine *Nature et Santé*, Vrije Universiteit *2009*

EXPÉRIENCE INDUSTRIELLE

Développement de ressources de langue pour le néerlandais, Proxem, Paris 2015

STAGES

6 mois, Laboratoire de Linguistique Formelle, CNRS, Université Paris Diderot 2015
Modélisation de l'acquisition des catégories de mots

2 mois, Laboratoire Modyco, CNRS, Université Paris-X 2014
Développement d'un classifieur de tweets autour d'événements culturels

1,5 mois et 2 semaines, Systran Traduction Automatique, Paris 2013 & 2014
Développement de ressources linguistiques pour le néerlandais

COMPÉTENCES INFORMATIQUES

Langages de programmation

· Python, C++, R, Java, OCaml, Perl, L^AT_EX

Applications

· R-Studio, Git, Eclipse, Weka, Unitex, Libre Office, Microsoft-Office, MySQL, Heurist, Prolific, Psychopy, Pavlovia

Systemes d'opération

· Linux, MacOS, Windows

LANGUES

Néerlandais : Native, Français & Anglais : Bilingue, Espagnol & Allemand : Intermédiaire

PARTICIPATION À DES COMITÉS DE LECTURE

Relectrice pour la revue *Language and Linguistics Compass* 2022

Relectrice pour la revue *Journal of Psycholinguistic Research (JOPR)* 2022

Relectrice pour *The 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies : Student Research Workshop (NAACL-HLT SRW 2021)* 2021

Relectrice pour la *Conference on Computational Linguistics (COLING)* 2020 & 2022

Relectrice pour la 12^{ième} édition de la *Language Resources and Evaluation Conference (LREC 2020)* 2020

Relectrice pour *The 2019 International Conference on Recent Advances in Natural Language Processing (RANLP 2019)* 2019

Membre du comité du programme du Workshop Cognitive Modeling and Computational Linguistics 2019 & 2020

Relectrice pour la session étudiants de la European Summer School of Language, Logic and Information 2017 & 2018

PUBLICATIONS

ARTICLES PUBLIÉS DANS DES REVUES À COMITÉ DE LECTURE

International

- Seminck, Olga, Philippe Gambette, Dominique Legallois, Thierry Poibeau. 2022. *The Evolution of the Idiolect over the Lifetime : A Quantitative and Qualitative Study of French 19th Century Literature*. Dans *Journal of Cultural Analytics*, 7 (3). <https://doi.org/10.22148/001c.37588>
- Brusini, Perrine, Olga Seminck (2ième premier auteur), Pascal Amsili, Anne Christophe. 2021. *The Acquisition of Noun and Verb Categories by Bootstrapping from a few Known Words : A Computational Model*. Dans *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2021.661479>
- Seminck, Olga, Philippe Gambette, Dominique Legallois, Thierry Poibeau. 2021. *The Corpus for Idiolectal Research (CIDRE)*. Dans *Journal of Open Humanities Data* 7 (15). <https://doi.org/10.5334/johd.42>
- Seminck, Olga. 2020. *The Use of Naturalistic Reading Corpora for the Study of Pronoun and Coreference Resolution*. Dans *Language and Linguistics Compass* 30 (3). <https://doi.org/10.1111/lnc3.12395>

CHAPITRES D'OUVRAGES

International

- Delaborde, Marine, Auph lie Ferreira, Lo c Grobol, Olga Seminck, Benjamin Fagard. Usages et perceptions du langage inclusif : des pratiques langagi res clivantes ? Accept    para tre dans *Language and Gender* (Fagard & Abrantes, eds.)

ARTICLES SOUMIS DANS DES ACTES DE CONF RENCES   COMIT  DE LECTURE

International

- Lasri, Karim, Olga Seminck, Alessandro Lenci, Thierry Poibeau. Accept    The 29th International Conference on Computational Linguistics (COLING). *Subject Verb Agreement Error Patterns in Meaningless Sentences : Humans vs. BERT*.

ARTICLES PUBLI S DANS DES ACTES DE CONF RENCES   COMIT  DE LECTURE

International

- Bulteau, Laurent, Philippe Gambette, Olga Seminck. 2022. *Reordering a tree according to an order on its leaves*. Dans les Actes de 33rd Annual Symposium on Combinatorial Pattern Matching (CPM 2022). Article No. 24 ; pages : 24 :1–24 :15. <https://doi.org/10.4230/LIPIcs.CPM.2022.24>
- Gambette, Philippe, Olga Seminck, Dominique Legallois, Thierry Poibeau. 2021. *Using and Evaluating Hierarchical Clustering Methods for Corpora with Chronological Order*. Dans les Actes de The 2021 Conference of the European Association for Digital Humanities (EADH).
- Seminck, Olga, Louise-Am lie Cougnon, Bernard Hanseeuw, C drick Fairon. 2020a *The Mind-it Corpus : a Longitudinal Corpus of Electronic Messages Written by Older Adults with Incipient Alzheimer's Disease and Clinically Normal Volunteers* Dans Proceedings de l'atelier Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments (RaPID-3). Marseille.

- Seminck, Olga & Pascal Amsili. 2018. *A Gold Anaphora Annotation Layer on an Eye Movement Corpus*. Dans Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC). Miyazaki. pages : 3518-3522. ISBN : 979-10-95546-00-9. <http://www.lrec-conf.org/proceedings/lrec2018/summaries/318.html>
- Seminck, Olga & Pascal Amsili. 2017. *A Computational Model of Human Preferences for Pronoun Resolution*. Dans Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia. pages 53-63. <https://aclanthology.coli.uni-saarland.de/papers/E17-4006/e17-4006>
- Amsili, Pascal & Olga Seminck. 2017a. *A Google-proof Collection of French Winograd Schemas*. Dans Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes (CORBON). Valencia. pages : 24-29. <http://www.aclweb.org/anthology/W17-1504>

National

- Wang, Xiaou, Olga Seminck, Pascal Amsili. 2022. *Investigating associative, switchable and negatable Winograd items on renewed French data sets*. Dans les Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale. Pages 136-143.
- Naguib, Marco, Marine Delaborde, Blandine Andrault, Anaïs Bekolo, Olga Seminck. 2022. *Romanciers et romancières du XIXème siècle : une étude automatique du genre sur le corpus GIRLS*. Dans Traitement Automatique des Langues Naturelles. pages : 66-77. ATALA.
- Seminck, Olga, Vincent Segonne, Pascal Amsili. 2019. *Modèles de langue appliqués aux schémas Winograd français*. Dans les actes de la 26ième Conférence sur le Traitement Automatique des Langues Naturelles (TALN). Toulouse. pages 343-350. https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_TALN-RECITAL-courts_CH_PFIA2019-2.pdf.
- Amsili, Pascal & Olga Seminck. 2017b. *Schémas Winograd en français : une étude statistique et comportementale*. Dans les actes de la 24ième Conférence sur le Traitement Automatique des Langues Naturelles (TALN). Orléans. pages : 28-35. http://taln2017.cnrs.fr/wp-content/uploads/2017/06/actes_TALN_2017-vol2.pdf.
- Seminck, Olga. 2016. *Un modèle simple du coût cognitif de la résolution des anaphores*. Dans Actes de la 18ième Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL). Paris. pages : 66-79. <https://jep-taln2016.limsi.fr/actes/Actes%20JTR-2016/V03-RECITAL.pdf> Prix du meilleur article RECITAL.

COMMUNICATIONS SANS ACTES À COMITÉ DE SÉLECTION PRÉSENTÉES À DES CONFÉRENCES

International

- Grunspan, Claude, Frédérique Mélanie-Becquet, Jean Barré, Laurette Chardon, Ioana Galleron, Marco Naguib, Clément Plancq, Olga Seminck, Thierry Poibeau. 2022. *Event annotation for literary corpora analysis*. Dans Digital Humanities 2022, ADHO, Tokyo, Japon.
- Galleron, Ioana, Rosario Arias, Javier Fernández-Cruz, Roxana Patraş, Olga Seminck, Frédérique Mélanie-Becquet. 2022. *What's in a preface? Sentiment analysis of liminal matter in ELTEC collections*. Dans Distant Reading for European Literary History - Final Action Event. Cracovie, Pologne.
- Delaborde, Marine, Auphémie Ferreira, Loïc Grobol, Gabrielle Le Tallec, Benjamin Fagard, Olga Seminck. 2021. *Usages et perception du langage inclusif : des pratiques langagières clivantes ?* Dans Entre féminin et masculin – langue(s) et société. Lisbonne, Portugal.
- Seminck, Olga, Philippe Gambette, Dominique Legallois, Thierry Poibeau. 2021. *The Corpus for Idiolectal Research (CIDRE)*. Dans The 2021 Conference of the European Association for Digital Humanities (EADH).
- Seminck, Olga, Louise-Amélie Cougnon, Bernard Hanseeuw, Cédric Fairon. 2020b *The Use of Patient's Electronic Message History for the Automatic Preclinical Detection of Alzheimer's Disease*. Dans Alzheimer's Association International Conference. Amsterdam.
- Seminck, Olga & Pascal Amsili. 2019. *A Cost Metric for Pronoun Resolution : Uncertainty Increases Processing Cost*. Dans Rational Approaches in Language Science. Saarbrücken. page : 27. http://www.sfb1102.uni-saarland.de/?page_id=57
- Seminck, Olga & Pascal Amsili. 2018. *The Time Course of Anaphora Resolution in Natural Text Reading*. Dans Linguistic Evidence. Tübingen. pages : 121- 123. https://uni-tuebingen.de/index.php?eID=tx_securedownloads&p=103205&u=0&g=0&t=1544982799&hash=bcd9a4783cd9bc89059c812bcc7b9ee2e5e02bf1&file=/fileadmin/Uni_Tuebingen/SFB/SFB_833/LE_2018/Booklet_komplett.pdf

- Seminck, Olga & Pascal Amsili. 2016. *A Cognitively Plausible Model for Anaphora Resolution*. Dans Architectures and Mechanisms of Human Language Processing. Bilbao. page : 155. <https://www.bcbl.eu/events/files/galeria/Book%20of%20abstracts%20AMLaP%202016.pdf>

National

- Seminck, Olga & Pascal Amsili. 2017. *Predicting Processing Cost of Anaphora Resolution*. Dans 39. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS). Saarbrücken. page : 97. http://dgfs2017.uni-saarland.de/wordpress/abstracts/DGfS_2017_booklet.pdf

RESOURCES DISPONIBLES EN LIGNE

Seminck, Olga, Blandine Andrault, Anaïs Bekolo, Marine Delaborde, Marco Naguib. *The Gender Identification Resource for Literature and Style (GIRLS)* <https://www.ortolang.fr/market/corpora/girls>.

Seminck, Olga, Philippe Gambette, Dominique Legallois, Thierry Poibeau. *The Corpus for Idiolectal Research (CIDRE)* <https://www.ortolang.fr/market/corpora/cidre/>.

Seminck, Olga. *Bootstrapping Model* https://github.com/oseminck/bootstrapping_model

Seminck, Olga & Pascal Amsili. *Anaphoric Pronouns and their Antecedent in the Dundee Eye-Tracking Corpus*. <http://www.llf.cnrs.fr/apadec>

Amsili, Pascal & Olga Seminck. *French Winograd Schemas*. <http://www.llf.cnrs.fr/winograd-fr>

VULGARISATION SCIENTIFIQUE

Seminck, Olga, Louise-Amélie Cougnon, Bernard Hanseeuw, Cédric Fairon. 2020c. *Les messages électroniques pour dépister la maladie d'Alzheimer*. Louvain Médical.

Seminck, Olga & Louise-Amélie Cougnon. 2019. Site-web sur le projet Mind-it avec des explications destinées au grand public (participants et intéressés). Université catholique de Louvain. <https://www.uclouvain-mindit.be>

C'est quoi ta thèse ? 2019. Vidéo d'explication des résultats de thèse en 3 minutes. Laboratoire d'excellence Emperical Foundations of Linguistics. <http://www.labex-efl.com/wordpress/2019/02/21/cest-quoi-ta-these-olga-seminck-modeles-cognitifs-et-computationnels-de-la-resolution-des-pronoms/>

Coste, Christophe & Olga Seminck. 2017. *Don't blame it on algorithms : what they really are and how they can fuel progress in the life sciences*. The Conversation. <https://theconversation.com/dont-blame-it-on-algorithms-what-they-really-are-and-how-they-can-fuel-progress-in-the-life-sciences-79471>

COMMUNICATIONS PRÉSENTÉES DANS DES JOURNÉES D'ÉTUDE SANS COMITÉ DE SÉLECTION

The Use of Electronic Messages for the Detection of Alzheimer's Disease : the MIND-IT Project, 2019
Institute of Neuroscience Young Researcher Day, Université catholique de Louvain

Peut-on détecter la maladie d'Alzheimer en utilisant les conversations électroniques ?, 2019
Journée sur la recherche et l'innovation sociale dans l'économie wallonne, Cercle de Wallonie, Namur

A Computational Model of Human Preferences for Pronoun Resolution, Journée des doctorants 2017
du Laboratoire de Linguistique Formelle, Université Paris Diderot

A Cognitively Plausible Model for Anaphora Resolution, Retraite de l'école doctorale Frontières du Vivant, 2016
Centre de Recherches Interdisciplinaires

A Computational Model for Processing Cost in Anaphora/Coreference Resolution, Anaphora Workshop, 2015
Université Paris Diderot

CONFÉRENCES DANS DES SÉMINAIRES

- Ordre des feuilles d'un arbre et évolution de l'idiolecte de romanciers et romancières du XIXe siècle, Laboratoire d'informatique Gaspard-Monge 2022
- Discussion de l'article *What BERT is not : Lessons from a New Suite of Psycholinguistic Diagnostics for Language Models* ensemble avec David Kletz, Groupe de travail Traitement Automatique des Langues du Laboratoire Lattice 2022
- L'évolution de l'idiolecte, Lattice Groupe de lecture Humanités Numériques, <https://www.lattice.cnrs.fr/recherche/groupe-de-lecture-humanites-numeriques/> 2021
- Cognition et TAL, Séminaire du Centre de traitement automatique du langage, Université catholique de Louvain, <https://uclouvain.be/fr/instituts-recherche/ilc/cental/seminaires-du-cental.html> 2019
- Peut-on Utiliser les conversations Électroniques pour la detection automatique de la maladie d'Alzheimer ?, Séminaire Alzheimer de l'Institute of Neuroscience, Université catholique de Louvain 2019
- Modèles cognitifs et computationnels de la résolution des pronoms, Séminaire de Linguistique Computationnelle, Sorbonne Université, http://lejeunegael.fr/seminaire_LC.html 2018
- Cognitive Computational Models of Pronoun Resolution, Séminaire du Computational Linguistics Group, Universität des Saarlandes 2018
- The Time Course of Pronoun Resolution in Natural Text Reading, Séminaire de Linguistique Expérimentale, Université Paris Diderot 2018
- Schémas Winograd en français : une étude statistique et comportementale, Séminaire LingLunch, Université Paris Diderot, <http://www.linguist.univ-paris-diderot.fr/linglunch/archives> 2018
- Predicting Processing Cost of Anaphora Resolution, Séminaire Lattice, Université Sorbonne Nouvelle 2017
- Présentation d'un article de Andrew Kehler et Hannah Rohde : A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation, Séminaire de Linguistique Expérimentale, Université Paris Diderot 2016
- A Computational Model for Processing Cost in Anaphora Resolution, Séminaire Interdisciplinary Friday, Centre de Recherches Interdisciplinaires, <https://cri-paris.org/news/interdisciplinary-friday-nov-27-olga-seminck-j-boussier-david-drummond/> 2015