

# Pascal Denis

PhD

Equipe MAGNET, Inria Lille–Nord Europe  
40, avenue Halley  
59650 Villeneuve d'Ascq, France  
☎ +33 3 59 35 87 24  
✉ [pascal.denis@inria.fr](mailto:pascal.denis@inria.fr)  
🌐 <http://researchers.lille.inria.fr/pdenis/>

## Employment

- 2012–present **Tenured Research Scientist (CRCN)**, *Inria*, Lille, France.  
Researcher and deputy-head of the MAGNET research group, a joint team between Inria Lille–Nord Europe and University of Lille Computer Science department (CRISTAL), focused on research in Machine Learning and Natural Language Processing.
- 2008–2012 **Tenured Research Scientist (CR2, CR1)**, *Inria*, Paris, France.  
Member of the ALPAGE Natural Language Processing research group at Inria Paris and University Paris-Diderot. Promoted Research Scientist 1<sup>st</sup> class (CR1) in 2010.
- 2010 **Visiting Scholar**, *University of Oxford*, Oxford, UK.  
Associate member of the Computational Linguistics Group, Department of Computer Science (ComLab).
- 2007–2008 **Natural Language Scientist**, *Powerset Inc. (now Microsoft)*, San Francisco, USA.  
Member of the Knowledge Resources team, in charge of the development of highly scalable software components for named-entity recognition and tracking, and of their integration in the Powerset Web search engine.
- 2000–2007 **Research & Teaching Assistant**, *The University of Texas at Austin*, Austin, USA.  
Research and teaching in linguistics and computer science.
- 2005 **Visiting Research Fellow**, *IMS*, Stuttgart, Germany.  
Five month research stay at the Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, funded by German Academic Exchange Service (DAAD).

## Education

- 2007 **Ph.D. in Computational Linguistics**, *The University of Texas at Austin*, USA.  
◦ Dissertation: *New Learning Models for Robust Reference Resolution*  
◦ Advisors: Jason Baldridge and Nicholas Asher  
◦ Committee: Laurence Danlos, Andrew Kehler, Jonas Kuhn, Raymond Mooney  
◦ GPA: 4.0/4.0
- 2019 **M.A. in Statistics**, *Université Libre de Bruxelles*, Brussels, Belgium.  
◦ Major: theoretical statistics  
◦ Honors: *Summa cum laude*
- 2000 **M.A. in Artificial Intelligence**, *Université Paul Sabatier*, Toulouse, France.
- 1999 **M.Sc. in Cognitive Science**, *University of Edinburgh*, Edinburgh, UK.  
◦ Major: Natural Language Processing
- 1998 **B.A. in Romance Philology**, *Université Libre de Bruxelles*, Brussels, Belgium.  
◦ Honors: *Magna cum laude*
- 1996 **B.A. in Journalism & Communication Studies**, *Université Libre de Bruxelles*, Brussels, Belgium.  
◦ Honors: *Cum laude*

## Honors and Awards

- 2018 **Distinguished Senior Program Committee Member**, *IJCAI-ECAI 2018 conference*.  
Distinction awarded by IJCAI-ECAI Program Chair for exceptional reviewing contribution
- 2017 **Prime d'Encadrement Doctoral et de Recherche (PEDR)**, *Inria*.  
National award of scientific excellence in research and PhD supervision
- 2013 **Best Paper Award**, *TALN 2013*, Les Sables-d'Olonne, France.  
Award given to the author(s) of the best paper presented at the annual, French NLP conference

- 2007 **Best Student Paper Award**, *NAACL-HLT 2007*, Rochester, NY, USA.  
Award given to the author(s) of the best paper written solely or primarily by a student appearing in the annual NAACL conference proceedings.
- 2005 **Homer Lindsey Bruce Fellowship**, *The University of Texas at Austin*, USA.  
Nine-month fellowship awarded in recognition of major achievements since entering Graduate School. Awarded only to the strongest applicants for the University of Texas Continuing University Fellowships.
- 2005 **Young Researcher Fellowship**, *German Academic Exchange Services (DAAD)*.  
Five-month scholarship awarded to outstanding graduate and post-doctoral students for conducting study and/or research at a German university.

---

## Research Grants and Projects

### As Principal Investigator

- 2020-2023 **Improving Embeddings with Semantic Knowledge (IMPRESS)**, *Bilateral Inria-DFKI project*, Co-PI: Ivana Kruijff-Korbayová (DFKI), Budget: ~ 500k€.
- 2016-2021 **Graph-based Machine Learning for Linguistic Structure Prediction (GRASP)**, *National Research Agency (ANR) Early Career Grant*, France, Budget: ~ 250k€.

### As Project Member

- 2020-2023 **Spin and Bias in Language Analyzed in News and Texts (SLANT)**, *ANR International Collaboration Grant (PRCI)*, France-Luxembourg, PI: Ph. Muller (IRIT).  
Role: Local Coordinator
- 2016-2021 **LEarning GOod representations for natural language processing (LEGO)**, *Inria-USC*, France-USA, Inria Associate Team with Fei Sha's group at USC.
- 2011-2021 **Empirical Foundations of Linguistics (EFL)**, *Laboratoire d'Excellence (Labex)*, Université Sorbonne Paris Cité.  
Role: Associate Member
- 2016-2021 **Rethinking English Modal (REM)**, *ANR International Collaboration Grant (PRCI)*, France-Switzerland, PIs: I. Depraetere (Univ. Lille) & M. Hilpert (Univ. Neuchâtel).
- 2011-2017 **Strategic Conversation (STAC)**, *European Research Council (ERC) Advanced Grant*, PI: Nicholas Asher, CNRS Toulouse.
- 2014-2017 **Structuring Discourse in Multilingual Europe (TextLink)**, *EU COST Action*, PIs: L. Degand (UCL, Louvain) and B. Webber (University of Edinburgh).
- 2010-2012 **Dynamically Enriching Lexical Resources in Multilingual and Multimodal Applications (EdyLex)**, *ANR Collaborative Grant (PRCE)*, France, PI: Benoît Sagot, Inria Paris.
- 2009-2011 **Large coverage probabilistic syntactic parsing of French (SEQUOIA)**, *ANR Collaborative Grant (PRCE)*, France, PI: Alexis Nasr, Université de la Méditerranée.
- 2006-2008 **Extracting and Using Discourse Structure to Resolve Anaphoric Dependencies: Combining Logico-Semantic and Statistical Approaches**, *National Science Foundation (NSF)*, USA, PI: N. Asher, Co-PIs: J. Baldridge; J. Kuhn, The University of Texas at Austin.
- 2002-2003 **Linguistic Research on Text Interpretation**, *Morris Memorial Grant*, *New York Community Trust*, USA, PI: Carlota Smith, The University of Texas at Austin.

---

## Software Development Grants and Projects

- 2020-2023 **Improving Embeddings with Semantic Knowledge (IMPRESS)**, *Action de Développement Technologique (ADT)*, *Inria*, France.  
Role: Project Leader
- 2016-2018 **Magnet Word Embedding Toolkit (MANGOES)**, *Action de Développement Technologique (ADT)*, *Inria*, France.  
Role: Project Leader

2008-2009 **Semi-automatic and Collaborative Retrieval of Information Based on Ontologies (SCRIBO)**,  
*Pôle de compétitivité System@tic, France*, PI: Eric de la Clergerie, Inria Paris.  
Role: Project Member

---

## Publications

### Conference Articles

- 2020 Onkar Pandit, Pascal Denis, and Liva Ralaivola. "Integrating Knowledge Graph Embeddings to Improve Mention Representation for Bridging Anaphora Resolution". In: *Proceedings of the Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2020)*. Barcelona, Spain (Online).
- Mariana Vargas-Vieyra, Pascal Denis, and Liva Ralaivola. "Joint Learning of the Graph and the Data Representation for Graph-Based Semi-Supervised Learning". In: *Proceedings of the Workshop on Graph-Based Natural Language Processing (TextGraphs 2020)*. Barcelona, Spain (Online).
- 2019 Mathieu Dehouck and Pascal Denis. "Phylogenetic Multi-Lingual Dependency Parsing". In: *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)*. Minneapolis, Minnesota, USA.
- Bo Li, Mathieu Dehouck, and Pascal Denis. "Modal sense classification with task-specific context embeddings". In: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019)*. Bruges, Belgium.
- Onkar Pandit, Pascal Denis, and Liva Ralaivola. "Learning Rich Event Representations and Interactions for Temporal Relation Classification". In: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019)*. Bruges, Belgium.
- M. Vargas Vieyra, A. Bellet, and P. Denis. "Probabilistic End-to-End Graph-based Semi-Supervised Learning". In: *NeurIPS Bayesian Deep Learning Workshop*. Vancouver, Canada.
- 2018 Melissa Ailem et al. "A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images". In: *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2018)*. Brussels, Belgium.
- Mathieu Dehouck and Pascal Denis. "A Framework for Understanding the Role of Morphology in Universal Dependency Parsing". In: *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2018), Short Papers*. Brussels, Belgium.
- 2017 Mathieu Dehouck and Pascal Denis. "Delexicalized Word Embeddings for Cross-lingual Dependency Parsing". In: *Proceedings of the European Association for Computational Linguistics (EACL)*. Valencia, Spain.
- Pascal Denis and Liva Ralaivola. "Online Learning of Task-specific Word Representations with a Joint Biconvex Passive-Aggressive Algorithm". In: *Proceedings of the European Association for Computational Linguistics (EACL)*. Valencia, Spain.
- 2016 Chloé Braud and Pascal Denis. "Learning Connective-Based Word Representations for Implicit Discourse Relation Classification." In: *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2016)*. Austin, Texas.
- 2015 Chloé Braud and Pascal Denis. "Comparing Word Representations for Implicit Discourse Relation Classification." In: *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2015)*. Lisbon, Portugal.
- Emmanuel Lassalle and Pascal Denis. "Joint Anaphoricity Detection and Coreference Resolution with Constrained Latent Structures". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2015)*. Austin, Texas.

- 2014 Chloe Braud and Pascal Denis. "Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification". In: *Proceedings of the International Conference on Computational Linguistics (COLING 2014)*. Association for Computational Linguistics. Dublin, Ireland.
- David Chatel, Pascal Denis, and Marc Tommasi. "Fast Gaussian Pairwise Constrained Spectral Clustering". In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2014)*. Nancy, France.
- 2013 Chloé Braud and Pascal Denis. "Identification automatique des relations discursives "implicites" à partir de données annotées et de corpus bruts". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013)*. Vol. 1. Sables d'Olonne, France, pp. 104–117.
- Emmanuel Lassalle and Pascal Denis. "Apprentissage d'une hiérarchie de modèles à paires spécialisés pour la résolution de la coréférence". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013)*. Les Sables-d'Olonne, France.
- Emmanuel Lassalle and Pascal Denis. "Improving pairwise coreference models through feature space hierarchy learning". In: *Proceedings of the Annual meeting of the Association for Computational Linguistics (ACL 2013)*. Association for Computational Linguistics. Sofia, Bulgaria.
- Antoine Venant et al. "Expressivity and comparison of models of discourse structure". In: *Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Metz, France.
- 2012 Stergos Afantenos et al. "Developing a corpus of strategic conversation in The Settlers of Catan". In: *SeineDial 2012 - The 16th Workshop on the Semantics and Pragmatics of Dialogue*. Paris, France.
- Philippe Muller et al. "Constrained decoding for text-level discourse parsing". In: *Proceedings of the International Conference on Computational Linguistics (COLING 2012)*. Mumbai, India.
- 2011 André Bittar, Pascal Amsili, and Pascal Denis. "French TimeBank : un corpus de référence sur la temporalité en français". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2011)*. Vol. 1. Laboratoire d'Informatique de Robotique et de Microélectronique & Université de Montpellier 2. Montpellier, France, pp. 259–270.
- André Bittar et al. "French TimeBank: An ISO-TimeML Annotated Reference Corpus". In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2011), Short Papers*. Portland, Oregon, USA: Association for Computational Linguistics, pp. 130–134.
- Pascal Denis and Philippe Muller. "Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition". In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2011)*. Barcelone, Spain.
- Enrique Henestroza Anguiano and Pascal Denis. "FreDist: Automatic construction of distributional thesauri for French". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2011), Short Papers*. Montpellier, France, France, pp. 119–124.
- Emmanuel Lassalle and Pascal Denis. "Leveraging different meronymy discovery methods for bridging resolution in French". In: *DAARC 2011 - 8th Discourse Anaphora and Anaphor Resolution Colloquium*. Ed. by A. Branco I. Hendrickx S. Lalitha Devi and R. Mitkov. Vol. 7099. Lecture Notes in Computer Science. Faro, Portugal: Springer, pp. 35–46.
- 2010 Stergos Afantenos et al. "Learning Recursive Segments for Discourse Parsing". In: *Language Resources and Evaluation*. Ed. by Nicoletta Calzolari (Conference Chair) et al. La Valette, Malta: European Language Resources Association (ELRA).
- Marie Candito, Benoît Crabbé, and Pascal Denis. "Statistical French dependency parsing: treebank conversion and first results". In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)*. La Valletta, Malta: European Language Resources Association (ELRA), pp. 1840–1847.

- Marie Candito et al. "Benchmarking of Statistical Dependency Parsers for French". In: *Proceedings of the International Conference on Computational Linguistics (COLING 2010)*. 9 pages. Beijing, China, pp. 108–116.
- Pascal Denis and Philippe Muller. "Comparison of different algebras for inducing the temporal structure of texts". In: *Proceedings of the International Conference on Computational Linguistics (COLING 2010)*. Beijing, China, pp. 250–258.
- Pascal Denis and Benoît Sagot. "Exploitation d'une ressource lexicale pour la construction d'un étiqueteur morphosyntaxique état-de-l'art du français". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2010)*. Montréal, Canada.
- 2009 Marie Candito et al. "Analyse syntaxique du français : des constituants aux dépendances". In: *Actes de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2009)*. Senlis, France.
- Éric De La Clergerie et al. "Extracting and Visualizing Quotations from News Wires". In: *LTC 2009 - 4th Language and Technology Conference*. Ed. by Zygmunt Vetulani. Vol. 6562. Lecture Notes in Artificial Intelligence. Poznań, Poland: Springer, pp. 522–532.
- Pascal Denis and Benoît Sagot. "Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort". In: *Pacific Asia Conference on Language, Information and Computation*. Hong Kong, China.
- 2008 Pascal Denis and Jason Baldridge. "Specialized Models and Ranking for Coreference Resolution". In: *Empirical Methods on Natural Language Processing ((EMNLP 2007))*. Ed. by The Association for Computational Linguistics. Honolulu, Hawaiï, USA, pp. 660–669.
- 2007 Pascal Denis and Jason Baldridge. "A ranking approach to pronoun resolution". In: *International Joint Conferences on Artificial Intelligence*. Hyderabad, India.
- Pascal Denis and Jason Baldridge. "Joint determination of anaphoricity and coreference resolution using integer programming". In: *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007)*. Rochester, New York, USA, pp. 236–243.
- 2006 Nicholas Asher, Pascal Denis, and Brian Reese. "Names and pops and discourse structure". In: *Workshop on Constraints in Discourse*. Maynooth, Ireland.
- Nicholas Asher et al. "Evidentiality and intensionality: Two uses of reportative constructions in discourse". In: *Workshop on Constraints in Discourse Structure*. Maynooth, Ireland.
- Pascal Denis and Jonas Kuhn. "Applying an LFG-parser for coreference resolution: Experiments and analysis". In: *International Lexical Functional Grammar Conference*. Constance, Germany.
- 2005 Nicholas Asher and Pascal Denis. "Lexical ambiguity as type disjunction". In: *International Workshop on Generative Approaches to the Lexicon*. Ed. by Pierrette Bouillon and Kyoko Kanzaki. Genève, Switzerland, pp. 10–17.
- 2003 Pascal Denis. "A fresh look at goal motion constructions". In: *Chicago Linguistics Society*. Chicago, Illinois, USA.
- Pascal Denis, Jonas Kuhn, and Stephen Wechsler. "V-PP goal motion complexes in English: an HPSG account". In: *Proceedings of the ACL-SIGSEM workshop: The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*. Toulouse, France, pp. 121–132.
- Pascal Denis and Philippe Muller. "A semantics for temporally dependent referring expressions". In: *Colloque de Syntaxe et Semantique a Paris*. Ed. by Olivier Bonami and Patricia Cabredo Hofherr. Paris, France, pp. 45–62.

## Journal Articles

- 2021 Aurélien Bellet et al. "Pour plus de transparence dans l'analyse automatique des consultations ouvertes : leçons de la synthèse du Grand Débat National". In: *Statistique et Société*. to appear.
- 2018 Bert Cappelle, Pascal Denis, and Mikaela Keller. "Facing the facts of 'fake': a distributional semantics and corpus annotation approach". In: *Yearbook of the German Cognitive Linguistics Association* 6.9-42.
- 2014 Chloe Braud and Pascal Denis. "Identifier les relations discursives implicites en combinant données naturelles et données artificielles". In: *Traitement automatique des langues* 55.
- 2012 Pascal Denis and Benoît Sagot. "Coupling an annotated corpus and a lexicon for state-of-the-art POS tagging". In: *Language Resources and Evaluation* 46.4, pp. 721–736.
- 2009 Pascal Denis and Jason Baldridge. "Global joint models for coreference resolution and named entity classification". In: *Procesamiento del Lenguaje Natural* 42.
- 2005 Pascal Amsili, Pascal Denis, and Laurent Roussarie. "Anaphores abstraites en français: représentation formelle". In: *Traitement Automatique des Langues* 46.1, pp. 15–39.
- 2003 Laurent Prevot et al. "Une approche sémantique et rhétorique du dialogue. Un cas d'étude: l'explication d'itinéraires". In: *Traitement Automatique des Langues* 43.2, pp. 43–70.
- 2001 Joan Busquets and Pascal Denis. "L'ellipse modale en français: le cas de "pouvoir" et "devoir"". In: *Les Cahiers de Grammaire* 26, pp. 55–74.

### Editorial Work

- 2006 Pascal Denis et al., eds. *Proceedings of the 2004 Texas Linguistics Society Conference: Issues at the Semantics-Pragmatics Interface*. Cascadilla Proceedings Project.

## Selected Invited Talks

- *Scalab Seminar*, Université Lille 3 France, December 2016
- *STL Seminar*, Université Lille 3, France, February 2016
- *IMS Guest Lecture Series*, Universität Stuttgart, Germany, July 2015
- *LPL Seminar*, Université de Aix-Marseille, Aix-en-Provence May 2015
- *30 Minutes de Science*, Inria Lille–Nord Europe, May 2015
- *Computational Linguistics Group Seminar*, University of Oxford, August 2011
- *Computational Linguistics Colloquium*, Universität Heidelberg, July 2009
- *Talaris Seminar*, Loria, Nancy, May 2009
- *Alpage Seminar*, Université Paris 7, Paris, October 2008
- *Job talk*, Powerset Inc., San Francisco, CA, March 2007
- *Job talk*, Department of Linguistics, Indiana University, Bloomington, IN, February 2007
- *IMS Institutversammlung*, Universität Stuttgart, Germany, January 2006
- *Job talk*, Département de Linguistique, Université d'Ottawa, Canada, February 2006
- *Lattice Seminar*, Université Paris 7, Paris, June 2005.

## Advising and Mentoring

### PhD Supervision

- 2020–present **Priyansh Trivedi**, *Improving Linguistic Embeddings with Semantic Knowledge (working title)*, Computer Science Graduate School, Université de Nancy.  
Co-supervised with Philippe De Groote
- 2020–present **Gaurav Maheshwari**, *Learning Trustworthy Representations for Natural Language Processing (working title)*, Computer Science Graduate School, Université Lille.  
Co-supervised with Aurélien Bellet

- 2018–present **Mariana Vargas**, *Adaptive Graph Learning with Applications to Natural Language Processing (working title)*, Computer Science Graduate School, Université Lille.  
Co-supervised with Aurélien Bellet and Rémi Gilleron
- 2017–present **Onkar Pandit**, *Graph-based Semi-supervised Linguistic Structured Prediction (working title)*, Computer Science Graduate School, Université Lille.  
Co-supervised with Liva Ralaivola and Marc Tommasi
- 2015–2019 **Mathieu Dehouck**, *Multi-Lingual Dependency Parsing: Word Representation and Joint Training for Syntactic Analysis*, Computer Science Graduate School, Université de Lille, Defended on 20 May 2019; **2020 Best PhD Thesis Award, ATALA**.  
Co-supervised with Marc Tommasi
- 2012–2017 **David Chatel**, *Semi-supervised Clustering in Graphs*, Computer Science Graduate School, Université Lille, Defended on 27 November 2017.  
Co-supervised with Marc Tommasi
- 2012–2015 **Chloé Braud**, *Identification automatique des relations discursives implicites à partir de corpus annotés et de données brutes*, Linguistics Graduate School, Université Paris-Diderot, Defended on 18 December 2015; **2016 Best PhD Thesis Award, ATALA**.  
Co-supervised with Laurence Danlos
- 2011–2015 **Emmanuel Lassalle**, *Structured learning for coreference resolution: a joint approach*, Computer Science Graduate School, Université Paris-Diderot, Defended on 11 May 2015.  
Co-supervised with Laurence Danlos
- 2008–2010 **André Bittar**, *Building a TimeBank for French: A Reference Corpus Annotated According to the ISO-TimeML Standard*, Linguistics Graduate School, Université Paris-Diderot, Defended on 19 November 2010.  
Co-supervised with Laurence Danlos and Pascal Amsili

### Master Thesis Supervision

- 2020 **Imad Aouali**, *Detecting Semantic Shift Using Word Embeddings*, Ecole Centrale de Lille & Master Program in Applied Mathematics, Université de Lille.
- 2016 **Thibault Liétard**, *Metric Learning for Graph-based Coreference Resolution*, Master Program in CS and Telecommunication, ENS Rennes, (Co-supervised with A. Bellet).
- 2015 **Mathieu Dehouck**, *Graph-based semi-supervised dependency parsing*, MOCAD Master Program, Université de Lille.
- 2011 **Chloé Braud**, *Implicit discourse relation identification*, Master Program in Computational Linguistics, Université Paris-Diderot.
- 2010 **Emmanuel Lassalle**, *Meronymic relation extraction for bridging resolution*, Master Program in Computational Linguistics, Université Paris-Diderot.

### Engineer Supervision

- 2018-2019 **François Noyez**, Engineer at Inria Lille.  
Worked on consolidating the CoRTex library, including adding a French component. Funded by Inria and Orange Labs.
- 2016-2018 **Nathalie Wauquier**, Engineer at Inria Lille.  
Worked on ADT Project MANGOES. Funded by Inria Lille

### Internship Supervision

- 2017 **Juhi Tandon**, IIT Hyderabad, India, May-August 2017.  
Worked on multilingual graph-based syntactic dependency parsing.
- 2011 **Altaf Rahman**, University of Texas at Dallas, USA, June-August 2011.  
Worked at consolidating the CoRTex Python library.
- 2009 **Alexis Vanacker**, Ecole des Mines de Nancy, June-August 2009.  
Worked at adapting the phrasetective.org annotation tool to French.

## Teaching Experience



### As Main Instructor

- 2021 **Learning on Graphs and Networks**, *Master Program in Data Science*, Université de Lille, Graduate level.  
Spring 2021 (32h)
- 2017-2019 **Foundations in Machine Learning**, *Master Program in Computer Science*, Université de Lille, Graduate level.  
Spring 2017, 2018, 2019 (32h)
- 2012-2014 **Information Extraction (Machine Learning)**, *MOCAD Master Program*, Université de Lille, Graduate level.  
Fall 2012 (15h), Fall 2013 (24h)
- 2014-2016 **Machine Learning and Decision Under Uncertainty**, *MOCAD Master Program*, Université de Lille, Graduate level.  
Fall 2014, 2015 (30h)
- 2012-2014 **Information Extraction (Machine Learning)**, *MOCAD Master Program*, Université de Lille, Graduate level.  
Fall 2012 (15h), Fall 2013 (24h)
- 2014-2016 **Machine Learning and Decision Under Uncertainty**, *MOCAD Master Program*, Université de Lille, Graduate level.  
Fall 2014, 2015 (30h)
- 2012-2014 **Information Extraction (Machine Learning)**, *MOCAD Master Program*, Université de Lille, Graduate level.  
Fall 2012 (15h), Fall 2013 (24h)
- 2009-2012 **Natural Language Processing (Semantics)**, *Master Program in Computational Linguistics*, Université Paris-Diderot, Graduate level.  
Fall 2009 (12h), Fall 2010 (24h), Fall 2011 (24h)
- 2010 **Automated Reference Resolution**, *Journées de Sémantique et de Modélisation (JSM)*, Loria, Nancy, Invited outreach lecture.  
Fall 2010 (2h)

### As Teaching Assistant

- 2003 **Computational Linguistics I**, *The University of Texas at Austin*, Graduate level, Instructor: Jonas Kuhn.  
Spring 2013
- 2003 **Introduction to Cognitive Science**, *The University of Texas at Austin*, Undergraduate level, Instructor: Carlota Smith.  
Spring 2013
- 2002 **Introduction to Computational Linguistics**, *The University of Texas at Austin*, Undergraduate level, Instructor: Bernard Schwarz.  
Fall 2002
- 2000 **Language and the Internet**, *The University of Texas at Austin*, Undergraduate level, Instructor: Rajesh Bhatt.  
Fall 2000

### Language Instructor

- 2001-2003 **Introductory and advanced French language classes**, *Summer School in French Language and Literature*, Université Libre de Bruxelles.

---

## Service to the Community

### Conference Chairing and Senior PC Membership

- 2021 Program Co-Chair of TALN
- 2020 ACL Area Chair (Multidisciplinary track); TALN Senior PC
- 2018 IJCAI-ECAI Senior PC (Natural Language Processing)



- 2016 IJCAI Senior PC (Natural Language Processing)
- 2015 ACL Area Chair (Discourse, Coference, Pragmatics)
- 2011 Constraints in Discourse (CID) Co-chair

### Conference Program Committee

- 2021 NAACL
- 2020 COLING, EMNLP, First Workshop on Computational Approaches to Discourse (CODI)
- 2019 ACL, EMNLP
- 2018 ACL, CONLL, EMNLP, NAACL-HLT, NAACL CRAC workshop, NIPS
- 2017 CAp, EACL, IWCS
- 2016 ACL, CAp, EMNLP, NAACL
- 2015 CAp, EMNLP, IJCAI, NAACL
- 2014 ACL, COLING, EACL, EMNLP, Journée AFIA/ATALA: Langue, Apprentissage, et Fouille de Textes
- 2013 ACL, EACL, EMNLP, \*SEM
- 2012 ACL, AAI, COLING, EACL, TALN, ATALA Workshop on Discourse and NLP
- 2011 ACL, CID, IJCAI, EMNLP, TALN
- 2010 COLING, EMNLP, ATALA Workshop on CRFs, ATALA Workshop on Corpus Annotation
- 2009 ACL, EMNLP, IEEE-ICSC, NAACL/HLT Workshop on Integer Linear Programming for NLP
- 2008 CICLing, SuB, TLS XII
- 2007 ACL
- 2005 LSRL

### Journal Reviewing

- Computational Linguistics
- Transactions of the Association for Computational Linguistics (TACL) (standing reviewer: 2020-2023)
- Natural Language Engineering (NLE)
- Language Resources and Evaluation
- IJCAI-ECAI Surveys
- ACM Transactions on Speech and Language Processing (TSLP)
- ACM Computing Surveys
- Traitement Automatique des Langues (TAL)
- Journal of Linguistics
- Discours

### Conference & Workshop Organizing

- 2021 **PC Co-Chair and Organizer**, *TALN 2021*, June-July 2021, Lille, France.  
Annual French NLP conference
- 2014-2016 **PC Co-Chair and Organizer**, *Polaris Colloquium Series*, Lille, France.  
Monthly Guest Lecture series in Computer Science and Signal Processing, co-sponsored by Inria Lille–Nord Europe and University of Lille CRISTAL Lab
- 2016 **Organizer**, *Workshop on Argumentation Mining*, June 2016, MESHS, Lille, France.
- 2011 **Co-organizer**, *Constraints in Discourse (CID 2011)*, June 2011, Agay, France.
- 2010 **Co-organizer**, *TempEval-2: Evaluating Events, Time Expressions, and Temporal Relations*, July 2010, Uppsala, Sweden.
- 2004 **Co-chair**, *Texas Linguistic Society 8th Annual Conference (TLS VIII)*, May 2004, Austin, TX, USA.

### Grant Proposal Reviewing

- 2016 Reviewer for Flanders Research Foundation (FWO, Belgium), Lead Agency Project Call
- 2016, 2019 External Expert for the Brussels Institute for Research and Innovation (Innoviris, Belgium)

2013, 2018 Reviewer for the French National Research Agency (ANR)

### Hiring Committees

- 2018-present **National Committee for Scientific Research (CoNRS)**, *CNRS National Hiring and Evaluation Committee*, Section 34: Language Sciences.  
Elected member
- 2020 **Inria Lille Research Center**, *Search committee for Junior Researchers CRCN/IFSP*, May 2020.  
Vice-President (resigned)
- 2020 **Université Paris-Diderot, Linguistics Department**, Search for Assistant Professor in Computational Semantics, May 2020.  
External member
- 2018 **Inria Lille Research Center**, *Search committee for Junior Researchers CRCN*, May 2018.  
Jury Member
- 2016 **Université Lille 3, CS Department**, Search for Assistant Professor in Data Science, May 2016.  
External member
- 2012 **Centre National de la Recherche Scientifique (CNRS)**, Search for a Research Engineer in NLP, September 2012.  
External expert
- 2012 **Université de Marseille, CS Department**, Search for Assistant Professor in Machine Learning/NLP, May 2012.  
External member
- 2011 **Université de Marseille, CS Department**, Search for Assistant Professor in Machine Learning/NLP, May 2011.  
External member
- 2004–2005 **University of Texas at Austin, Linguistics Department**, *Search for Assistant Professor in Computational Linguistics*, Fall 2004.  
Student representative

### Evaluation Committees

- 2019 **High Council for Evaluation of Research and Higher Education (HCERES)**.  
Expert and CoNRS Representative
- 2015–2017 **Commission Emploi Recherche (CER)**, *Inria Lille*.  
Committee Member
- 2017-2018 **Commission du Développement Technologique (CDT)**, *Inria Lille*.  
Committee Member
- 2015 **ATALA PhD Award Committee**, May 2015.  
Committee Member

### PhD Defense Committees

- 2021 **Rémi Cardon**, Linguistics Department, Université de Lille, TBD.  
External examiner
- 2018 **Elena Knyazeva**, Computer Science Department, Université Paris-Saclay, Orsay, May 2018.  
Rapporteur
- 2015 **Juliette Conrath**, Computer Science Department, Université Paul Sabatier (UPS), Toulouse, December 2015.  
External examiner
- 2012 **Emili Sapena**, Computer Science Department, Universitat Politècnica De Catalunya (UPC), Barcelona, May 2012.  
External examiner

### PhD Supervisory Committees

- 2020-present **Zulpiye Yusupjiang**, Linguistics Department, Université Paris-Diderot, Supervisor: Jonathan Ginzburg.
- 2019-present **Farah Cherfaoui**, CS Department, Université de Marseille, Supervisors: Sandrine Anthoine and Liva Ralaivola.
- 2018-present **Rémi Cardon**, Linguistics Department, Université de Lille, Supervisor: Natalia Grabar.
- 2017-2020 **Loic Grobol**, Linguistics Department, Université Paris Sorbonne, Supervisors: Isabelle Tellier and Eric de la Clergerie.

### Miscellaneous

- 2010-present Administrator of Inria membership to Linguistic Data Consortium (LDC)

---

## Memberships in Professional Organizations

- The Association for Computational Linguistics (ACL)
- ACL's Special Interest Group on Natural Language Learning (SIGNLL)
- ACL's Special Interest Group on Semantics (SIGSEM)
- ACL's Special Interest Group on the Lexicon (SIGLEX)
- The French Association for Natural Language Processing (ATALA)
- CNRS Groupement de recherche (GDR) TAL

---

## Open Source Software Development

**MANGOES**, *Magnet Word Embedding Toolkit*, Python toolbox for constructing and evaluating word vector representations, <https://gitlab.inria.fr/magnet/mangoes/>.

Co-author and developer with Nathalie Vauquier and Rémi Gilleron

**CoRTex**, *Coreference Resolution Toolkit*, Python library for coreference resolution, <https://gforge.inria.fr/projects/cortex/>.

Original author and main developer

**MELT**, *Maximum-Entropy Lexicon-Enriched Tagger*, <http://lingwb.gforge.inria.fr/>, High precision lexically-enriched, discriminative POS tagger for French..

Co-author and developer with Benoît Sagot

**POLKA**, *Python Online Learning toolKit for lAanguage*, Python library for online learning, <https://gforge.inria.fr/projects/polka/>.

Original author and main developer

**Statgram**, Resources for the statistical dependency parsing of French, incl. preprocessing code and parsing models for MaltParser, MSTParser, Berkeley Parser, [http://alpage.inria.fr/statgram/frdep/fr\\_stat\\_dep\\_parsing.html](http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html).

Co-author and developer with M. Candito, B. Crabbé, F. Guérin, E. Henestroza Anguiano, J. Nivre, D. Seddah

**FreDist**, Python library for constructing distributional thesauri from raw texts, <https://gforge.inria.fr/projects/fredist/>.

Co-author and developer with E. Henestroza Anguiano

**SAPIENS**, *Newswire Quotation Extractor & Navigator*, <http://passage.inria.fr/SAPIENS/>.

Developer of the anaphora resolution module

---

## Technological Transfer

- 2018 **Orange Labs**, *Meylan*, 6-month collaboration with Orange Deskin team on developing coreference resolution systems for French Orange datasets.
- 2018 **Claim Assistance SAS**, *Pantin*, 6-month collaboration on argumentation mining and legal NLP.
- 2017-2018 **Start-up Wordzit**, *Lille*, Member of Scientific Advisory board.
- 2012 **Studio national des arts contemporains, Le Fresnoy**, *Tourcoing*, Collaboration with contemporary artist Véronique Bélard on project THIS IS MAJOR TOM TO GROUND CONTROL.



## References

Available on request.