

# Thierry Poibeau

---

---

## Main Research Areas

Natural Language Processing  
Machine learning  
Machine translation  
General linguistics  
Semantics  
Digital Humanities  
Epistemology and History of Sciences

---

## Affiliation and Personal Details

Affiliation **Laboratoire LATTICE**  
(Langues, Textes, Traitements informatiques, Cognition)  
UMR 8094 CNRS, ENS/PSL and U. Sorbonne nouvelle/USPC  
École normale supérieure  
45 rue d'Ulm  
75005 Paris  
Tel. +33 1 58 07 66 20

Mail [thierry.poibeau@ens.fr](mailto:thierry.poibeau@ens.fr)  
Web <http://www.lattice.cnrs.fr/Thierry-Poibeau/>  
Publications <https://cv.archives-ouvertes.fr/thierry-poibeau>

---

## Professional Experience

2012– **Head of the LATTICE Research Unit** (CNRS, École normale supérieure/PSL and Université Sorbonne nouvelle/USPC)

2011– **Directeur de Recherche (Director of Research)** in General Linguistics and Natural Language Processing at CNRS (UMR 8094)

2008– **Affiliated lecturer**, Dpt of Theoretical and Applied Linguistics, Univ. of Cambridge

2003–2009 **Chargé de Recherche (Research Fellow)** in computer science at CNRS

2002–2003 **Associated Professor at INALCO** (Institut National des Langues et Civilisations Orientales)

1998–2003 **Research engineer at Thales/TRT** (Thales Research and Technology, France)

---

## Education

2008 **Habilitation in computer science**, Université Paris-Nord (Paris 13)  
Automatic semantic analysis for information access

- 2002 **PhD in computer science**, Université Paris-Nord (Paris 13)  
Information extraction from texts
- 1994 **MA in theoretical linguistics and natural language processing**, Université Denis-Diderot (Paris 7)
- 1993 **MA in classics**, Université Paris-Sorbonne (Paris 4)

---

## Main teaching duties (2012–present)

- 2017– **Master in Digital Humanities** at École normale supérieure.  
Natural language processing for Digital Humanities (24 hours/year)
- 2016– **PhD course**, Paris Sciences et Lettres  
Manual and automatic text annotation (24 hours/year)
- 2013– **MPhil in linguistics**, University of Cambridge  
Lectures on Lexical semantics + “reading club” (12 hours/year)
- 2012–2014 **MA seminar**, École normale supérieure  
Computational linguistics for Digital humanities and literary text processing (24 hours/year)
- 2014 **PhD course**, Université Sorbonne nouvelle  
Introduction to natural language processing (24 hours)
- until 2015 **MA “Multilingual Engineering”**, INaLCO  
Information extraction (24 hours/year)

---

## Research Projects (2012–present)

- 2016– **ERC LEXICAL Project** (project participant within the univ. of Cambridge, under the supervision of Dr Anna Korhonen)  
Large-scale multilingual lexical acquisition
- 2015– **ERA-NET Project Atlantis** (head of the project for LATTICE)  
Modelling language emergence in a community of robots
- 2015– **Projet PSL Lakme** (head of the the project)  
Parsing morphologically-rich texts (Hebrew, Medieval French and Uralic languages)
- 2011–2019 **Labex TRANSFERS**  
“Excellence initiative” gathering all the humanities labs at École normale supérieure
- 2010–2012 **European FP7 STREP PANACEA Project**  
Linguistic resource induction from raw texts

---

## PhD students (2012–present)

- 2016– **KyungTae Lim**, ENS Ulm  
Natural language parsing for robot interaction
- 2014– **Miquel Cornudella Gaya**, ENS Ulm / Sony CLS Paris  
The Role of Intrinsic Motivation in Artificial Language Emergence
- 2014– **Jean-Christophe Pautre**, Bibliothèque nationale de France  
Semantic analysis of the Gallica collection

- 2014–2017 **Pablo Ruiz**, ENS Ulm  
Natural Language Processing for Digital Humanities
- 2011–2014 **Elisa Omodei**, ENS Ulm / ISC-PIF  
Modeling the evolution of scientific domains using large scientific text collections
- 2010–2015 **Pierre Marchal**, INaLCO  
Large scale linguistic resource induction from raw texts with an application to Japanese
- 2010–2014 **Zorana Ratkovic**, Université Sorbonne nouvelle / INRA  
Distributional textual analysis for biological natural language processing
- 2009–2013 **Mani Ezzat**, INaLCO  
Large scale linguistic resource induction from raw texts
- 2009–2013 **Yufan Guo**, University of Cambridge (co-supervised with A. Korhonen)  
Automatic Text Zoning in Scientific Discourse

---

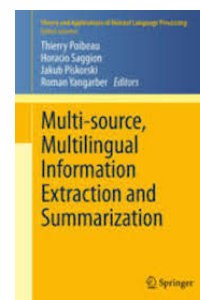
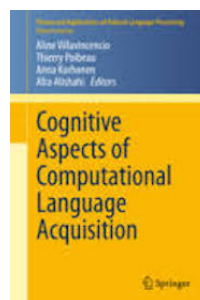
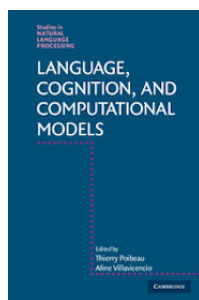
## Other Duties and Responsibilities

- 2012– Head of the LATTICE Research Units (20 permanent researchers and 20 PhD students)
- 2011– Member of the scientific committee of the labex Transfers (ENS, <http://transfers.ens.fr/>)
- Member of the editorial board of the journal TAL (2009–2016) and now of *Language and Linguistic Compass* (2016–)
- Reviewer for journals (e.g. JNLE – *Journal of Natural language Engineering*; RNTI – *Revue des Nouvelles Technologies de l'Information*, etc.); and conferences (e.g. ACL; IJCAI; ECAI; COLING; LREC; COGSCI; TALN; Lexis and Grammar, etc.)
- Organizer of various scientific events (e.g. TALN-2016 *Celtic Language Technology Workshop*; EMNLP-2015 *Workshop on Cognitive Aspects of Computational Language Learning*; EACL-2014 *Workshop on Cognitive Aspects of Computational Language Learning*; 2013 *Workshop on Language, Cognition and Computational Models*; EACL-2012 *Workshop on Computational Models of Language Acquisition and Loss*)

# Recent Publications (selection)

Only publications since 2012 are taken into account

## Books and Edited Books



- 2017 Thierry Poibeau. *Machine translation*. MIT Press (coll. Essential knowledge), Cambridge (USA), 300 pages, 2017.
- 2017 Thierry Poibeau et Aline Villavicencio (eds.). *Language, Cognition and Computational Models*. Cambridge University Press, Cambridge, env. 320 pages, 2017.
- 2013 Aline Villavicencio, Thierry Poibeau, Anna Korhonen, Afra Alishahi (eds.). *Cognitive Aspects of Computational Language Acquisition*. Springer, Berlin, 350 pages, 2013, ISBN : 978-3-642-31863-4. <hal-00782234>
- 2012 Thierry Poibeau, Horacio Saggion, Jakub Piskorski, Roman Yangarber (eds.). *Multi-source, Multilingual Information Extraction and Summarization*. Springer, Berlin, 330 pages, 2012, 978-3-642-28569-1. <hal-00782234>

## Articles in a peer-reviewed journal

- 2016 Thierry Poibeau. Traduire sans comprendre ? La place de la sémantique en traduction automatique. *Langages* n° 201, Armand Colin, 2016, pp. 77–90. <hal-01273768>
- 2014 Thierry Poibeau, Shravan Vasishth. Cognitive Issues in Natural Language Processing. In *Revue TAL (Traitement Automatique des Langues, numéro spécial "Traitement Automatique des Langues et Sciences Cognitives")* n° 55 (3), 2014, pp. 7–19. <10.1201/b21583-2>. <hal-01386307>
- Gilles Col, Thierry Poibeau. An instruction-based analysis of over. In *Language and Cognition*, Vol. 6 Issue 3, Cambridge University Press, 2014, pp. 370-407 <10.1017/l.angcog.2014.10>. <halshs-01282382>
- 2012 Gilles Col, Jeanne Aptekman, Stéphanie Girault, Thierry Poibeau. Gestalt Compositionality and Instruction-Based Meaning Construction. In *Cognitive Processing* n° 13 (2), Springer Verlag, 2012, pp.151-170. <10.1007/s10339-011-0431-y>. <hal-00782230>
- Barry Devereux, Nicholas Pilkington, Thierry Poibeau, Anna Korhonen. Towards unrestricted, large-scale acquisition of feature-based conceptual representations from corpus data. In *Research on Language and Computation* n° 7 (2-4), Springer Verlag, 2010, pp.137-170. <hal-00605539>

---

## Papers published in conference proceedings (selection)

- 2017 Ruiz Fabo, Pablo, Clara Martínez Cantón, Thierry Poibeau and Elena González-Blanco. (2017). Enjambment detection in a large diachronic corpus of Spanish sonnets. *LaTeCH-CLFL 2017 (Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature)*. Vancouver, Canada, 2017.
- 2016 Miquel Cornudella, Thierry Poibeau, Remi Van Trijp. The Role of Intrinsic Motivation in Artificial Language Emergence : a Case Study on Colour. *26th International Conference on Computational Linguistics (COLING)*, Osaka, Japan, pp.1646-1656, 2016. <hal-01422478>
- Pierre Marchal, Thierry Poibeau. A Continuum-based Model of Lexical Acquisition. *Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, Konya, Turkey, 2016. <hal-01349563>
- Pablo Ruiz, Clément Plancq, Thierry Poibeau. More than Word Cooccurrence : Exploring Support and Opposition in International Climate Negotiations with Semantic Parsing. *The 10th Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia. pp.1902-1912, 2016. <hal-01326692>
- Estelle Tieberghien, Frédérique Mélanie-Becquet, Pablo Ruiz Fabo, Thierry Poibeau, Melissa Terras. Mapping the Bentham Corpus. *Digital Humanities (DH2016)*, Krakow, Poland. 2016. <hal-01378029>
- 2015 Pablo Ruiz, Thierry Poibeau, Frédérique Mélanie. ELCO3 : Entity Linking with Corpus Coherence Combining Open Source Annotators. *Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT)*, Denver, USA. Companion volume (demonstrations), 2015. <hal-01173969>
- Pablo Ruiz, Thierry Poibeau. Combining Open Source Annotators for Entity Linking through Weighted Voting. *Joint Conference on Lexical and Computational Semantics (\*SEM)*, Denver, USA. pp.211-215, 2015. <hal-01173967>
- 2014 Elisa Omodei, Jean-Philippe Cointet, Thierry Poibeau. Mapping the Natural Language Processing Domain : Experiments using the ACL Anthology. *Ninth International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, pp.2972-2979, 2014. <hal-01056147>
- Thierry Poibeau. Optimality Theory as a Framework for Lexical Acquisition. *15th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, Nepal. 2014. <hal-00996763>
- 2013 Tim Van de Cruys, Thierry Poibeau, Anna Korhonen. A Tensor-based Factorization Model of Semantic Compositionality. *Conference of the North American Chapter of the Association of Computational Linguistics (HTL-NAACL)*, Atlanta, USA. pp.1142-1151, 2013. <hal-00997334>
- 2012 Tim Van de Cruys, Laura Rimell, Thierry Poibeau, Anna Korhonen. Multiway Tensor Factorization for Unsupervised Lexical Acquisition. *24th International Conference on Computational Linguistics (COLING)*, Mumbai, Inde. pp.2703-2720, 2012. <hal-00783711>
- Elisa Omodei, Thierry Poibeau, Jean-Philippe Cointet. Multi-Level Modeling of Quotation Families Morphogenesis. *ASE/IEEE International Conference on Social Computing (SOCIAL-COM)*, Amsterdam, Pays-Bas. ASE/IEEE, 2012. <hal-00783699>