Paul Lerner

Phone : +336 37 61 35 05
Email : lerner@isir.upmc.fr
LinkedIn : paul-lerner-466997a8
GitHub : github.com/PaulLerner
Website : https://paullerner.github.io/

Education

MCF Qualification in section 27 (Computer Science)

Qualification to apply to *Maître de conférences* positions (Assistant Professor) in section 27 (Computer science) of French National Council of Universities, valid until 2028 inclusive. Reviewers : Sylvain Castagnos (Université de Lorraine) and Thierry Delot (Université Polytechnique Hauts-de-France).

Paris-Saclay University, CNRS, LISN (ex-LIMSI) Orsay, France

Ph.D. in Computer Science, Knowledge-based Visual Question Answering about Named Entities 2020–2023

Ph.D. advised by Olivier Ferret (Université Paris-Saclay, CEA, List) and Camille Guinaudeau (Université Paris-Saclay, CNRS, LISN).

Jury members :

Pierre Zweigenbaum (Université Paris-Saclay, CNRS, LISN)	President
Josiane Mothe (IRIT, CNRS, Université Toulouse Jean-Jaurès)	Reviewer & Examiner
Philippe Mulhem (Université Grenoble Alpes, CNRS, LIG)	Reviewer & Examiner
Michel Crucianu (CEDRIC-CNAM)	Examiner
Ewa Kijak (University of Rennes, Inria, IRISA)	Examiner

Summer schools attended during my Ph.D. :

— ALPS 2021 : Natural Language Processing and Speech Processing

— LxMLS 2023 : Machine Learning and Natural Language Processing

Paris Descartes University Paris, France

M.S. in Artificial Intelligence, with high honors 2018–2019

Done as a double degree with ESILV.

M2 internship : *Parkinson's Disease Diagnosis based on Handwriting*, advised by Laurence Likforman-Sulem (Télécom ParisTech).

ESILV Courbevoie, France

Engineer in Computer Science 2014–2019

M1 internship : *Behavioral Strategies during Human-Machine Interaction*, advised by Beatrice Biancardi and Catherine Pelachaud (ISIR, Sorbonne Université).

Czech Technical University Prague, Czech Republic

B.S. in Computer Science (exchange with ESILV) 2016

Experience

Sorbonne University, CNRS, ISIR Paris, France

Postdoctoral researcher – Political biases of Large Language Models 2024–2026

Postdoc advised by François Yvon and Benjamin Piwowarski within MLIA team (Machine Learning for Information Access) at ISIR.

Keywords : Automatic Summarization, Machine Translation, Multilingual Large Language Model, Political Bias, Alignment, Multiculturalism.

Sorbonne University, CNRS, ISIR Paris, France
Postdoctoral researcher – Machine Translation of Scientific Neologisms 2023–2024

Postdoc advised by François Yvon within MLIA team at ISIR.

Keywords : Neologism, Terminological Variation, Morphology, Machine Translation, Multilingual Large Language Model, In-Context Learning.

Paris-Saclay University, CNRS, LISN (ex-LIMSI) Orsay, France
Ph.D. Student – Knowledge-based Visual Question Answering about Named Entities 2020–2023

Ph.D. advised by Olivier Ferret (Université Paris-Saclay, CEA, List) and Camille Guinaudeau (Université Paris-Saclay, CNRS, LISN), within TLP team (Spoken Language Processing) at LISN.

Keywords : Visual Question Answering, Multimodal Information Retrieval, Representation Learning, Named Entities, Pretraining

Paris-Saclay University, CNRS, LIMSI Orsay, France
Research Engineer – Multimodal Speaker Diarization October 2019–2020

Advised by Hervé Bredin and Camille Guinaudeau within TLP team at LIMSI.

Keywords : Multi-party Dialogues, Speaker Diarization, Named Entity Recognition, Named Entity Disambiguation, Forced Alignment, Active Learning, Multimodal

Télécom ParisTech Paris, France
Research Intern – Parkinson's Disease Diagnosis based on Handwriting March–September 2019

Advised by Laurence Likforman-Sulem within S2A team (Signal, Statistics and Machine Learning) at Télécom ParisTech.

Keywords : Parkinson's Disease, Deep Learning, Clinical Decision Support System

Sorbonne University (ex-Pierre and Marie Curie), CNRS, ISIR Paris, France
Engineer Intern – Behavioral Strategies during Human-Machine Interaction April–September 2018

Advised by Beatrice Biancardi and Catherine Pelachaud within PIRoS team at ISIR.

Keywords : Human-Machine Interaction, Reinforcement Learning, Conversational Agent

Advising

Joanna Radoła – Code Switching Generation Paris, France
M2 Internship – Sorbonne University, CNRS, ISIR February–August 2025

Internship co-advised by François Yvon.

Keywords : Code Switching, Multilingual Large Language Model

Salem Messoud – Multimodal Rerankig Orsay, France
M2 Internship – Paris-Saclay University, CNRS, LISN March–September 2022

Internship co-advised by Olivier Ferret and Camille Guinaudeau.

Keywords : Multimodal Information Retrieval, Reranking, Visual Question Answering, Representation Learning, Named Entities

Teaching

Teacher at Aivancity Cachan, France
Natural Language Processing 2024–2025

Design of classes and practical works. Design of a homework and exam per semester. Materials : https://paullerner.github.io/aivancity_nlp/

Content : Zipf's law, Distributional Semantics, Skip-gram, Word Embeddings, Recurrent Neural Networks, Attention Mechanism, Transformer, Large Language Models, Pretraining, Fine-tuning, Alignement (RLHF/DPO), Decoding, In-Context Learning, Evaluation Metrics, Ethics and Biases.

Teacher Assistant at ENSAE — Palaiseau, France
Deep Learning : Models and Optimization — 2023–2025

Assistant of Kevin Scaman (2023-2024) then Olivier Koch (2024-2025). Materials : https://kscaman.github.io/teaching/2023_ENSAE_DL.html

Content : Multi-layer Perceptron, Multi-class Classification (Cross Entropy), Convolutional Neural Network, Computer Vision, Variational Auto Encoder.

Teacher Assistant at ENSAE — Palaiseau, France
Machine Learning for Natural Language Processing — 2023–2025

Advising of projects and correction of project reports. Assistant of Christopher Kermorvant. Materials : https://github.com/Deep-NLP-Course

Content : Bag of Words and Classification, Word Embeddings and Analogies, Language Models and Decoding.

Teacher Assistant at Paris-Saclay University — Orsay, France
Introduction to Machine Learning — 2021–2023

Introduction to Machine Learning and Natural Language Processing. Advising of projects and examiner of project defenses. Monitoring of exams and correction of exams. Assistant of François Landes et Kim Gerdes. Materials : https://gitlab.inria.fr/flandes/ias

Content : Gradient Descent, Perceptron, Principal Component Analysis, Maximum Likelihood Estimation, Naive Bayes Classification, K-means, TF-IDF, Overfitting and Generalization, Pre-processing and Encoding.

Teacher Assistant at Polytech Paris-Saclay — Orsay, France
Imperative programming — 2020–2023

Monitoring of practical works and exams. Correction of practical works and homeworks. Assistant of Frédéric Voisin.

Content : Binary representation, Types, Loops and Conditions, Functions, Algorithms, Recursion.

Teacher Assistant at Polytech Paris-Saclay — Orsay, France
Software development — 2020–2023

Monitoring of practical works and exams. Correction of practical works and homeworks. Assistant of Joël Falcou.

Content : C++ standard library, input/output flows, pointers and references, operator and function overloading, struct, template, unit tests

Service

Reviewer

Year	Journal	International Conference	National Conference
2024		EMNLP, ACMMM, ACL, ICMR	JEP-TALN
2023	Pattern Recognition	ICMR	RECITAL-RJCRI
2022		ACMMM	

Fixed-term workers' representative at LIMSI's board — Orsay, France
Paris-Saclay University, CNRS, LIMSI — Octobre 2019–2020

Participation to the boards monthly meetings.

Publications

International Journals

[20] B. Biancardi, M. Mancini, **P. Lerner** et C. Pelachaud, « Managing an Agent's Self-Presentational Strategies During an Interaction », *Frontiers in Robotics and AI*, t. 6, p. 16, 2019, **Impact factor : 3.4**, Long, ISSN : 2296-9144.

National Journals

[16] **P. Lerner**, S. Messoud, O. Ferret, C. Guinaudeau, H. Le Borgne, R. Besançon, J. G. Moreno et J. Lovón Melgarejo, « Un jeu de données pour répondre à des questions visuelles à propos d'entités nommées », *Traitement Automatique des Langues*, t. 63, n° 2, p. 15-39, 2022, Long.

International Conferences

[1] M. Mancini, B. Biancardi, S. Dermouche, **P. Lerner** et C. Pelachaud, « Managing Agent's Impression Based on User's Engagement Detection », in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, sér. IVA '19, **Rang B**, Court, Poster, 3 pages, Paris, France : Association for Computing Machinery, 2019, p. 209-211, ISBN : 9781450366724.

[6] **P. Lerner**, O. Ferret et C. Guinaudeau, « Multimodal Inverse Cloze Task for Knowledge-Based Visual Question Answering », in *Advances in Information Retrieval (ECIR 2023)*, **Rang A**, Long, Présentation, Cham : Springer Nature Switzerland, 2023, p. 569-587, ISBN : 978-3-031-28244-7.

[7] **P. Lerner**, O. Ferret et C. Guinaudeau, « Cross-modal Retrieval for Knowledge-based Visual Question Answering », in *Advances in Information Retrieval (ECIR 2024)*, **Rang A**, Long, Présentation, Cham : Springer Nature Switzerland, 2024, p. 421-438, ISBN : 978-3-031-56027-9.

[12] **P. Lerner** et F. Yvon, « Unlike "Likely", "Unlike" is Unlikely : BPE-based Segmentation hurts Morphological Derivations in LLMs », in *Proceedings of the 31st International Conference on Computational Linguistics*, **Rang B**, Court, 5 pages, International Committee on Computational Linguistics, 2025.

[13] **P. Lerner** et F. Yvon, « Towards the Machine Translation of Scientific Neologisms », in *Proceedings of the 31st International Conference on Computational Linguistics*, **Rang B**, Long, 9 pages, International Committee on Computational Linguistics, 2025.

[18] **P. Lerner**, O. Ferret, C. Guinaudeau, H. Le Borgne, R. Besançon, J. G. Moreno et J. Lovón Melgarejo, « ViQuAE, a Dataset for Knowledge-based Visual Question Answering about Named Entities », in *Proceedings of The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, sér. SIGIR'22, **Rang A***, Long, Poster, New York, NY, USA : Association for Computing Machinery, 2022, p. 3108-3120.

[19] **P. Lerner**, J. Bergoënd, C. Guinaudeau, H. Bredin, B. Maurice, S. Lefevre, M. Bouteiller, A. Berhe, L. Galmant, R. Yin et C. Barras, « Bazinga ! A Dataset for Multi-Party Dialogues Structuring », in *Proceedings of the Language Resources and Evaluation Conference*, **Rang B**, Long, Poster, Marseille, France : European Language Resources Association, 2022, p. 3434-3441.

National Conferences

[3] **P. Lerner**, O. Ferret et C. Guinaudeau, « Recherche cross-modale pour répondre à des questions visuelles », in *18e Conférence en Recherche d'Information et Applications*, H. Zargayouna, éd., **Rang C**, Long, Présentation, Paris, France : ATALA, 2023, p. 74-92.

[4] **P. Lerner** et F. Yvon, « Vers la traduction automatique des néologismes scientifiques », French, in *Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position*, M. Balaguer, N. Bendahman, L.-M. Ho-dac, J. Mauclair, J. G Moreno et J. Pinquier, éd., **Rang C**, Long, Présentation, Toulouse, France : ATALA et AFPC, juill. 2024, p. 245-261.

[5] **P. Lerner**, O. Ferret, C. Guinaudeau, H. Le Borgne, R. Besançon, J. G. Moreno et J. Lovón Melgarejo, « Un jeu de données pour répondre à des questions visuelles à propos d'entités nommées en utilisant des bases de connaissances », in *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) 2022.*, **Rang C**, Court, Présentation, Avignon, France : ATALA, 2022, p. 434-444.

[17] **P. Lerner** et L. Likforman-Sulem, « Classification of Online Handwriting Time Series for Parkinson's Disease Diagnosis using Deep Learning », in *Proceedings of the 4th Junior Conference on Data Science and Engineering (JDSE) (non-archival)*, Court, Poster, 2019, p. 3.

International Workshops

[2] **P. Lerner** et C. Grouin, « INCLURE : a Dataset and Toolkit for Inclusive French Translation », in *Proceedings of the 17th Workshop on Building and Using Comparable Corpora (BUCC) @ LREC-COLING 2024*, P. Zweigenbaum, R. Rapp et S. Sharoff, éd., Long, Présentation, 10 pages, Torino, Italia : ELRA et ICCL, mai 2024, p. 59-68.

Invited Seminars

[8] **P. Lerner**, « Towards Machine Translation of Scientific Neologisms », in *IRIT seminars*, Toulouse, France, 2024.

[14] **P. Lerner**, « Automatic Data Annotation and Webly Supervised Visual Question Answering », in *Knowledge-Enhanced Information Retrieval workshop (KEIR @ ECIR 2024)*, Glasgow, Scotland, 2024.

[15] **P. Lerner**, « Morphological Competence of LLMs : Applied to Translation of Scientific Neologisms », in *ChangeLing (CMU) seminars*, online, 2025.

Software and datasets

[9] **P. Lerner**, *ViQuAE*, version v4.0.0-alpha, 16 jan. 2024.

[10] **P. Lerner**, *Inclure*, version v0.1.0, 7 avr. 2024.

[11] **P. Lerner**, `neott`, version v1.2.0, 17 déc. 2024.

Programming Skills

- **Deep learning :** PyTorch
- **Languages :** Python, C, C++, C#, Java
- **Python libraries :** Faiss, Transformers, spaCy, Scikit-learn, NumPy, Matplotlib, Seaborn
- **Computer clusters** : slurm
- **Databases :** SPARQL, Elasticsearch

Languages

- **French :** native
- **English :** fluent
- **Spanish :** elementary