

Paul Lerner

Email : lerner@tuta.io
LinkedIn : [paul-lerner-466997a8](https://www.linkedin.com/in/paul-lerner-466997a8)
GitHub : github.com/PaulLerner
Site web : <https://paullerner.github.io/>



ÉTUDES

Université Paris-Saclay, CNRS, LISN (ex-LIMSI)

Orsay, France

Thèse en Informatique, Répondre aux questions visuelles à propos d'entités nommées

2020–2023

Ma thèse a été préparée sous la direction d'Olivier Ferret (Université Paris-Saclay, CEA, List) et le co-encadrement de Camille Guinaudeau (Université Paris-Saclay, CNRS, LISN).

Membres du jury :

Pierre Zweigenbaum (Université Paris-Saclay, CNRS, LISN)	Président
Josiane Mothe (IRIT, CNRS, Université Toulouse Jean-Jaurès)	Rapporteur & Examinatrice
Philippe Mulhem (Université Grenoble Alpes, CNRS, LIG)	Rapporteur & Examinateur
Michel Crucianu (CEDRIC-CNAM)	Examinateur
Ewa Kijak (Université de Rennes, Inria, IRISA)	Examinatrice

Écoles d'été suivies dans ce cadre :

- ALPS 2021 : traitement automatique des langues et de la parole
- LxMLS 2023 : apprentissage automatique, en particulier appliqué au traitement des langues

Université Paris Descartes

Paris, France

Master Sciences, Technologies, Santé, spécialité Intelligence Artificielle, mention bien

2018–2019

Ce master a été préparé dans le cadre d'un double diplôme avec l'ESILV.

Stage de fin d'études : *Deep learning for Parkinson's Disease diagnosis based on handwriting examination*, encadré par Laurence Likforman-Sulem (Télécom ParisTech).

Ce master couvre de façon large divers champs de l'Intelligence Artificielle : apprentissage automatique et fouille de texte mais aussi systèmes multi-agents.

ESILV

Courbevoie, France

Ingénieur en Informatique

2014–2019

Stage effectué en 1^{re} année d'école (M1) : *Agent's Self-Presentational Strategies During an Interaction*, encadré par Beatrice Biancardi et Catherine Pelachaud (ISIR, Sorbonne Université).

L'ESILV est une école d'ingénieur généraliste post-bac (avec prépa intégrée). J'ai commencé à m'orienter vers l'informatique dès la première année (L3) et j'ai choisi la spécialité informatique en deuxième année (M1). L'ESILV fait partie d'un pôle universitaire aux côtés d'une école de management et de multimédia (design web) et met l'accent sur l'interdisciplinarité.

Czech Technical University

Prague, République Tchèque

Licence en informatique (échange avec l'ESILV)

2016

J'ai bénéficié du programme ERASMUS pour effectuer un échange d'un semestre à la CTU à Prague où j'ai étudié la théorie des graphes et la programmation en C. Par ailleurs, j'ai ainsi découvert un nouveau système d'enseignement.

Année	Matière	Établissement	Niveau	TD	TP
2023-2024	<i>Deep Learning : Models and Optimization</i>	ENSAE	M2 (Ingé 3)	–	6
2023-2024	<i>Machine Learning for Natural Language Processing</i>	ENSAE	M2 (Ingé 3)	–	9
2022-2023	Introduction à l'apprentissage statistique	UFR Sciences Paris-Saclay	L3	–	24
2022-2023	Bases du développement logiciel	Polytech Paris-Saclay	L3 (Ingé 1)	24	–
2022-2023	Programmation Impérative	Polytech Paris-Saclay	L1 (Prépa 1)	14	10
2021-2022	Introduction à l'apprentissage statistique	UFR Sciences Paris-Saclay	L3	–	24
2021-2022	Bases du développement logiciel	Polytech Paris-Saclay	L3 (Ingé 1)	24	–
2021-2022	Programmation Impérative	Polytech Paris-Saclay	L1 (Prépa 1)	14	10
2020-2021	Bases du développement logiciel	Polytech Paris-Saclay	L3 (Ingé 1)	22	–
2020-2021	Programmation Impérative	Polytech Paris-Saclay	L1 (Prépa 1)	16	16

Total éq. TD : 213h

Chargé de TP à l'ENSAE (vacataire) Palaiseau, France
Deep Learning : Models and Optimization 2023–2024

Responsable : Kevin Scaman. Support : https://kscaman.github.io/teaching/2023_ENSAE_DL.html
 Notions abordées : perceptron multi-couches, classification multi-classes (entropie croisée), réseau de neurones convolutionnel, traitement d'images, auto-encodeur variationnel.

Chargé de TP à l'ENSAE (vacataire) Palaiseau, France
Machine Learning for Natural Language Processing 2023–2024

Responsable : Christopher Kermorvant. Support : <https://github.com/Deep-NLP-Course>
 Notions abordées : Sac de mots et classification, plongements lexicaux (*word embedding*) et analogies, modèles de langues et génération.

Chargé de TD à l'UFR Sciences, Université Paris-Saclay (moniteur) Orsay, France
 Introduction à l'apprentissage statistique 2021–2023

Introduction à l'apprentissage automatique et au traitement automatique des langues. Encadrement et soutenance de projets ainsi que surveillance et correction d'examens. Responsables : François Landes et Kim Gerdes. Support : <https://gitlab.inria.fr/flandes/ias>
 Notions abordées : Descente de gradient, Perceptron, Analyse en composantes principales, Estimateur du maximum de vraisemblance, Classification naïve bayésienne, K-moyennes, TF-IDF, Surapprentissage et généralisation, Pré-traitement et encodage. Bibliothèques Python : NumPy et Scikit-learn.

Chargé de TD à Polytech Paris-Saclay (moniteur) Orsay, France
 Programmation Impérative 2020–2023

Surveillance et correction de TP notés et de devoirs maisons ainsi que surveillance d'examens. Responsable : Frédéric Voisin.
 Notions abordées : Représentations binaires, types, boucles et conditions, fonctions, algorithmes, récursivité.

Chargé de TD à Polytech Paris-Saclay (moniteur) Orsay, France
 Bases du développement logiciel 2020–2023

Surveillance et correction de TP notés et de devoirs maisons ainsi que surveillance d'examens. Responsable : Joël Falcou.
 Notions abordées : Bibliothèque standard du C++, flux entrée/sortie, pointeurs et références, surcharge d'opérateurs et de fonctions, struct, template, tests unitaires.

Traduction automatique des néologismes scientifiques

Paris, France

Postdoc, Sorbonne Université, CNRS, ISIR

2023–2024

Postdoc en collaboration avec François Yvon au sein de l'équipe MLIA de l'ISIR.

La recherche scientifique découvre et invente continuellement de nouveaux concepts qui sont alors désignés par de nouveaux termes, des néologismes ou néonymes dans ce contexte. Puisque les publications se font très majoritairement en anglais, il convient de traduire fidèlement ces termes dans d'autres langues, comme le français, tout en évitant une multiplication d'anglicismes. Nous proposons d'exploiter la définition du terme afin de le traduire plus fidèlement. Pour ce faire, nous explorons les capacités de modèles de langues multilingues, qui parviennent à traduire des néologismes scientifiques dans une certaine mesure. Nous montrons notamment qu'ils utilisent souvent des procédés morphosyntaxiques appropriés mais sont limités par la segmentation en unités sous-lexicales et biaisés par la fréquence d'occurrences des termes ainsi que par des similarités de surface entre l'anglais et le français (un article en cours de relecture).

Répondre aux questions visuelles à propos d'entités nommées

Orsay, France

Thèse en Informatique, Université Paris-Saclay, CNRS, LISN (ex-LIMSI)

2020–2023

Ma thèse a été préparée sous la direction d'Olivier Ferret (CEA List) et le co-encadrement de Camille Guinaudeau (LISN), au sein de l'équipe Traitement du Langage Parlé (TLP) du LISN.

Ma thèse se positionne à l'intersection de plusieurs domaines de recherche, le traitement automatique des langues, la Recherche d'Information (RI) et la vision par ordinateur, qui se sont unifiés autour des méthodes d'apprentissage de représentation et de pré-entraînement. Dans ce contexte, nous avons défini et étudié une nouvelle tâche multimodale : répondre aux questions visuelles à propos d'entités nommées (KVQAE). Dans ce cadre, nous nous sommes particulièrement intéressé aux interactions cross-modales et aux différentes façons de représenter les entités nommées. Nous avons également été attentifs aux données utilisées pour entraîner mais surtout évaluer les systèmes de question-réponse à travers différentes métriques.

Plus précisément, nous avons proposé à cet effet un jeu de données, le premier de KVQAE comprenant divers types d'entités. Nous avons également défini un cadre expérimental pour traiter la KVQAE en deux étapes grâce à une base de connaissances non-structurée et avons identifié la RI comme principal verrou de la KVQAE, en particulier pour les questions à propos d'entités non-personnes [6], [9]. Ce travail a été étendu dans le cadre du stage de M2 de Salem Messoud, que j'ai encadré avec Camille et Olivier, qui portait sur le réordonnement multimodal de la RI initiale [2].

Afin d'améliorer l'étape de RI, nous avons étudié différentes méthodes de fusion multimodale, lesquelles sont pré-entraînées à travers une tâche originale : l'*Inverse Cloze Task* multimodale [4].

Nous avons trouvé que ces modèles exploitaient une interaction cross-modale que nous n'avions pas considéré à l'origine, et qui permettrait de traiter l'hétérogénéité des représentations visuelles des entités nommées. Ces résultats ont été renforcés par une étude du modèle CLIP qui permet de modéliser cette interaction cross-modale directement [3], [8]. Ces expériences ont été menées tout en restant attentif aux biais présents dans le jeu de données ou les métriques d'évaluation, notamment les biais textuels qui affectent toute tâche multimodale.

Identification du locuteur multimodale

Orsay, France

Ingénieur de recherche (CDD), Université Paris-Saclay, CNRS, LIMSI

Octobre 2019–2020

Ce travail a été encadré par Hervé Bredin et Camille Guinaudeau, au sein de l'équipe Traitement du Langage Parlé (TLP) du LIMSI.

Les séries télévisées sont remplies de dialogues naturels et multipartites qui constituent un défi pour de nombreuses tâches de traitement automatique des langues et de la parole, telles que l'identification du locuteur et la reconnaissance d'entités nommées. Ces tâches, habituellement traités de façon mo-

nomodale, peuvent bénéficier de la multimodalité des vidéos. Par ailleurs, de nombreuses ressources externes, telles que la transcription des séries télévisées, sont disponibles et permettent une annotation semi-automatique, notamment à travers un alignement forcé de l’audio vers la transcription. Le corpus résultant, *Bazinga!*, comprend plus de 400 heures de parole et plus de 8 millions de tokens, dont plus de 500 000 tokens annotés avec le locuteur, l’allocutaire et des informations sur les liens entre les entités [5]. Nos expériences sur ce corpus démontrent la difficulté de transférer des modèles entraînés sur des corpus mono-locuteurs ou des textes écrits, pour l’identification du locuteur, la restauration de ponctuation et la reconnaissance d’entités nommées.

D’autres travaux non publiés portaient sur l’apprentissage actif et d’identification du locuteur multimodale. Ces deux méthodes se sont avérées moins efficaces que l’alignement forcé décrit dans [5].

Diagnostic de la maladie de Parkinson d’après un examen manuscrit

Paris, France

Stage de fin d’études, Télécom ParisTech

Mars–Septembre 2019

Ce stage était encadré par Laurence Likforman-Sulem au sein de l’équipe Signal, Statistique et Apprentissage (S2A) de Télécom ParisTech.

La maladie de Parkinson est la deuxième maladie neurodégénérative la plus répandue. Plusieurs chercheurs ont tenté de fournir un système d’aide à la décision clinique qui confirmerait ou remettrait en question le diagnostic d’un neurologue sur la base d’un examen manuscrit peu coûteux et non invasif, étant donné que de nombreux symptômes de la maladie de Parkinson sont observables à travers l’écriture. Nous poursuivons leur travail en explorant les techniques d’apprentissage de représentation [10].

Stratégies comportementales pendant une interaction homme-machine

Paris, France

Stage, Sorbonne Université, CNRS, ISIR

Avril–Septembre 2018

Ce stage était encadré par Beatrice Biancardi et Catherine Pelachaud au sein de l’équipe PIROS à l’ISIR.

Dans le cadre du doctorat de Beatrice Biancardi, nous avons cherché à concevoir un agent conversationnel (avatar virtuel) capable de gérer l’impression qu’il fait sur son interlocuteur humain. L’agent utilise différentes stratégies comportementales pour être perçu comme chaleureux et/ou compétent. J’ai mis en œuvre un algorithme d’apprentissage par renforcement qui utilise l’engagement de l’utilisateur dans l’interaction comme récompense [7]. Pour tester nos hypothèses, nous avons conçu et réalisé une expérience à la Cité des sciences et de l’industrie, à Paris, afin d’évaluer le système [1].

ACTIVITÉS D’INTÉRÊT COLLECTIF

Relecteur de revue

Année	Revue
2023	Pattern Recognition

Comité de programme

Année	Conférence internationale	Conférence nationale
2024	ACL, ICMR	JEP-TALN
2023	ICMR	RECITAL-RJCRI
2022	ACMMM	

Représentant des CDD au conseil du laboratoire

Université Paris-Saclay, CNRS, LIMSI

Orsay, France

Octobre 2019–2020

Participation aux réunions mensuelles.

Type de publication	Nombre
Revue internationale avec comité de lecture	1
Revue nationale avec comité de lecture	1
Conférences internationales avec comité de lecture	5
Conférences nationales avec comité de lecture	3
Ateliers internationaux avec comité de lecture	1
Séminaires invités	1
<i>Total</i>	12

Les trois travaux que j'aimerais présenter sont [3], [4], [6]. J'ai principalement conçu et implémenté les approches développées dans ces trois travaux, dans le cadre de ma thèse encadrée par Olivier Ferret et Camille Guinaudeau, comme décrit plus haut.

Revue internationale avec comité de lecture

- [1] 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. DOI : 10.3389/frobt.2019.00093.

Revue nationale avec comité de lecture

- [2] **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. adresse : <https://aclanthology.org/2022.tal-2.0>.

Conférences internationales avec comité de lecture

- [3] **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. DOI : 10.1007/978-3-031-56027-9_26.
- [4] **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. DOI : 10.1007/978-3-031-28244-7_36.
- [5] **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. adresse : <https://aclanthology.org/2022.lrec-1.367>.
- [6] **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. DOI : 10.1145/3477495.3531753.

- [7] 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, Paris, France : Association for Computing Machinery, 2019, p. 209-211, ISBN : 9781450366724. DOI : 10.1145/3308532.3329442.

Conférences nationales avec comité de lecture

- [8] **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. adresse : <https://aclanthology.org/2023.jeptalnrecital-coria.5>.
- [9] **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. adresse : <https://aclanthology.org/2022.jeptalnrecital-taln.43/>.
- [10] **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.

Ateliers internationaux avec comité de lecture

- [11] **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) at LREC 2024*, Long, Présentation, 2024. adresse : <https://hal.science/hal-04531938>.

Séminaires invités

- [12] **P. Lerner**, « Automatic Data Annotation and Webly Supervised Visual Question Answering », in *Knowledge-Enhanced Information Retrieval workshop (KEIR @ ECIR 2024)*, Glasgow, Scotland, 2024. adresse : <https://keirworkshop.github.io/>.

COMPÉTENCES EN PROGRAMMATION

- **Apprentissage de représentation** : PyTorch
- **Langages** : Python, C, C++, C#, Java
- **Bibliothèques Python** : Faiss, Transformers, spaCy, Scikit-learn, NumPy, Matplotlib, Seaborn
- **Clusters** : slurm
- **Bases de données** : SPARQL, Elasticsearch

LANGUES

- **Français** : langue maternelle
- **Anglais** : courant
- **Espagnol** : scolaire