

Julien BEZANÇON

Doctorant en TAL

SORBONNE UNIVERSITÉ LABORATOIRE STIH

UNITÉ DE SERVICE CERES

+33 6 28 78 82 88

julienbezancon@gmail.com

<https://julienbez.github.io>

PARCOURS ACADÉMIQUE

Mission d'enseignement (doctorat)	2022 - 2025
SORBONNE UNIVERSITÉ (3 ans)	
<i>Cours de base de données, d'algorithmique et d'ingénierie des langues</i>	
Stagiaire en entreprise	2022
MALAKOFF HUMANIS (6 mois)	
<i>Création d'un moteur de recherche, stage de fin de master</i>	
Stagiaire de recherche	2021
SORBONNE UNIVERSITÉ, COMÉDIE FRANÇAISE (2 mois)	
<i>Retranscription et annotation des Registres de la Comédie Française</i>	
Stagiaire de recherche	2020
SORBONNE UNIVERSITÉ, CERES (2 mois)	
<i>Analyse automatique des Mazarinades (XVII siècle)</i>	

DIPLÔMES UNIVERSITAIRES

Doctorat mention <i>mathématiques, informatique et applications</i>	2022 - 2025
SORBONNE UNIVERSITÉ, LABORATOIRE STIH, CERES	
<i>Traitement automatique du phénomène de défigement d'expressions multimots</i> , sous la direction de Gilles Siouffi, Gaël Lejeune et Antoine Gautier	
Master Traitement automatique des langues	2020 - 2022
SORBONNE UNIVERSITÉ	
mémoire : <i>Création d'un corpus de stéréotypes du français et évaluation des biais des modèles de langue existants</i> , sous la direction de Karèn Fort, Aurélié Névéol et Yoann Dupont	
Licence sciences du langage spécialité informatique	2017 - 2020
SORBONNE UNIVERSITÉ	

ENSEIGNEMENTS

Les volumes sont renseignés en heures équivalent TD (HETD).

Intitulé	Public	Année	Volume
Algorithme et structure de données	L2	2024	TD : 10 / CM : 0
Base de données	L3	2023-24	TD : 48 / CM : 0
Ingénierie des langues	L2	2023-25	TD : 120 / CM : 15
Dictionnaire et néologie	M1	2025	TD : 1.5 / CM : 1

PUBLICATIONS

Conférences internationales avec comité de lecture

- Julien Bezançon and Gaël Lejeune. For a fistful of puns: Evaluating a puns in multiword expressions identification algorithm without dedicated dataset. In *Findings of the 2025 Conference on Empirical Methods in Natural Language Processing (main: 22.16 acceptance rate, findings: 17.35 acceptance rate)*, 2025
- Karen Fort, Laura Alonso Alemany, Luciana Benotti, Julien Bezançon, Claudia Borg, Marthese Borg, Yongjian Chen, Fanny Duce, Yoann Dupont, Guido Ivetta, Zhijian Li, Margot Mieskes, Marco Naguib, Yuyan Qian, Matteo Radaelli, Wolfgang S. Schmeisser-Nieto, Emma R. Schulz, Thiziri Saci, Sarah Saidi, Javier Torroba Marchante, Shilin Xie, Sergio E. Zanutto, and Aurélie Névél. Your stereotypical mileage may vary: Practical challenges of evaluating biases in multiple languages and cultural contexts. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17764–17769, Torino, Italia, mai 2024. ELRA and ICCL
- Aurélie Névél, Yoann Dupont, Julien Bezançon, and Karën Fort. French CrowS-Pairs: Extending a challenge dataset for measuring social bias in masked language models to a language other than English. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8521–8531, Dublin, Ireland, mai 2022. Association for Computational Linguistics

Conférences nationales avec comité de lecture

- Julien Bezançon and Gaël Lejeune. Reconnaissance de défigements dans des tweets en français par des mesures de similarité sur des alignements textuels. In *30e Conférence sur le Traitement Automatique des Langues Naturelles*, pages 56–67, Paris, France, 2023. ATALA
- Aurélie Névél, Yoann Dupont, Julien Bezançon, and Karën Fort. French CrowS-Pairs: Extension à une langue autre que l'anglais d'un corpus de mesure des biais sociétaux dans les modèles de langue masqués (French CrowS-Pairs : Extending a challenge dataset for measuring social bias in masked language models to a language other than English). In *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, pages 355–364, Avignon, France, juin 2022. ATALA

Ateliers internationaux avec comité de lecture

- Julien Bezançon, Félix Alié, Antoine Gautier, Marceau Hernandez, and Gaël Lejeune. Forbidden fruit is the sweetest: An annotated tweet corpus for french unfrozen idioms identification. In Amir Zeldes and Ines Rehbein, editors, *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX)*, Vienna, Austria, juillet 2025. Association for Computational Linguistics
- Julien Bezançon, Rimane Karam, and Gaël Lejeune. Lost in variation: An unsupervised methodology for mining lexico-syntactic patterns in middle Arabic texts. In *Proceedings of the 4th Workshop on Arabic Corpus Linguistics (WACL-4)*, pages 25–37, Abu Dhabi, UAE, janvier 2025. Association for Computational Linguistics
- Karën Fort, Aurélie Névél, Yoann Dupont, and Julien Bezançon. Use of a Citizen Science Platform for the Creation of a Language Resource to Study Bias in Language Models for French: a case study. In *2nd LREC Workshop on Novel Incentives in Data Collection from People*, Marseille, France, juin 2022

Ateliers nationaux avec comité de lecture

- Julien Bezançon, Toufik Boubehziz, Corina Chutaux, Oumaima Zine, Laurie Acensio, Andrea Briglia, Ibtihel Ben Ltaifa, Nour El Houda Ben Chaabene, Caroline Koudoro-Parfait, and Gaël Lejeune. SPQR@Deft2023: Similarité Sorbonne Pour les Systèmes de Question Réponse. In *30e Conférence sur le Traitement Automatique des Langues Naturelles*, pages 11–22, Paris, France, 2023. ATALA

Conférences nationales sans comité de lecture

- Julien Bezançon. Cas d'Étude du défigement en tal : Brisons la glace. In *Approches interdisciplinaires des unités phraséologiques (UP) dans les langues du monde*, pages 19–20, Paris, France, 2025
- Julien Bezançon, Marceau Hernandez, Félix Alié, and Gaël Lejeune. « défricher des lettres » annotation de défigements dans des tweets avec le défricheur. In *Linguistique Informatique, Formelle et de Terrain 2024*, Orléans, France, 2024. GDR LIFT

AUTRES ACTIVITÉS LIÉES À LA RECHERCHE

Mentorat d'étudiant pour ACL 2025

Relecture d'un article scientifique étudiant avant soumission

Relecteur 2025

- ACL 2025, relecteur de second niveau
- TALN 2025, relecteur de second niveau

Présentation à la journée Sorbonne Actuelle 2025

SORBONNE UNIVERSITÉ

Intervenant invité, 14 juin 2025

Séminaire doctoral du CERES 2022 - 2025

SORBONNE UNIVERSITÉ

Membre du comité d'organisation

Présentations au séminaire Linguistique Computationnelle 2023 - 2025

SORBONNE UNIVERSITÉ

- *Lost in Variation : fouille de motifs lexico-syntaxiques dans des textes en Moyen Arabe*, 21 novembre 2025
- *Reconnaissance de défigements dans des tweets en français par des mesures de similarité sur des alignements textuels*, 11 mai 2023

Participations au Hackathon annuel CERES 2023 - 2024

SORBONNE UNIVERSITÉ

- *Découverte automatique de défigements dans des titres de films avec des approches d'apprentissage machine non-supervisées*, juin 2024
- *Étiquetage d'un corpus en Moyen Arabe avec des approches état de l'art*, juin 2023

Présentation à l'atelier mensuel CERES 2023

SORBONNE UNIVERSITÉ, CERES

Voyant Tools : introduction et présentation des fonctionnalités de l'outil, 7 février 2023

BOURSES

Bourse de thèse CERES, SORBONNE UNIVERSITÉ

2022

OUTILS ET RESSOURCES DÉVELOPPÉS

ASMR

ASMR (*Align, Segment, Match and Rank*) est un algorithme d'identification et d'extraction automatique de défigements candidats depuis des textes. Il s'agit d'une approche empirique basée sur des alignements de séquence et des mesures de similarité. Il a été conçu pour s'exécuter rapidement et avec peu de ressources. Dépôt GITHUB : <https://github.com/JulienBez/ASMR>

Le Défricheur

Le Défricheur est une plateforme d'annotation gamifiée, créée afin de détecter la présence de défigements dans des tweets. Les annotateurs jouent le rôle d'agents dont le but est de lutter contre la propagation de jeux de mots dans des expressions multi-mots en les identifiant et en les signalant. Voir publications éponymes pour plus d'informations. Lien : <https://ceres.huma-num.fr/defricheur/>

COMPÉTENCES INFORMATIQUES

Langages maîtrisés	: PYTHON, SQL, HTML/CSS, BASH, L ^A T _E X.
Autres langages connus	: C, JAVA.
Outils de TAL	: AntConc, Gate, Unitex, TXM.
Bases de données	: Neo4J, MongoDB, Protégé.
Moteurs de recherche	: Solr, Elasticsearch, Kibana.
Autres outils	: R, Praat, Github.

RÉFÉRENCES

Gaël Lejeune, gael.lejeune@sorbonne-universite.fr

Karën Fort, karen.fort@loria.fr

Aurélie Névéol, aurelie.neveol@lisn.fr