

Fanny Ducl

<https://fannyducel.github.io>
fanny.ducel@lisn.upsaclay.fr

Formation

Master Langue et Informatique

2021 – 2023

Sorbonne Université

Paris, France

- M2 : Major de promotion. Moyenne provisoire de 17,8/20. Mémoire sur l'Évaluation des biais stéréotypés dans les modèles de langues auto-régressifs : état de l'art et exploration de techniques extrinsèques
- M1 : Major de promotion. Moyenne annuelle de 17,35/20. Mémoire sur l'Analyse des *claims* dans les articles de Traitement Automatique des Langues à l'aide d'une méthode par apprentissage non supervisé

Licence Sciences du Langage, spécialité Informatique - mention Très Bien

2018 – 2021

Sorbonne Université

Paris, France

Expériences en recherche

Stage de Recherche (M2 - 35h/semaine)

Mars 2023 - Septembre 2023

LORLA, encadrée par Karěn Fort et Aurélie Névéol

Nancy, France

- Établissement d'un état de l'art sur les biais stéréotypés dans les modèles de langues auto-régressifs avec une approche critique
- Expériences de manipulation de modèles de langues masqués et auto-régressifs pour l'évaluation des biais, reproduction et adaptation d'expériences de limitation et quantification de biais
- Participation au projet de recherche international MultiCrowsPairs, constitué de 20 autres chercheur-es et étudiant-es de 8 pays différents

Assistante de Recherche (CDD à temps partiel - 17h/semaine)

Septembre 2022

CERES, Sorbonne Université, supervisée par Gaël Lejeune

Paris, France

- Participation à un projet de cartographie de chercheurs et chercheuses selon leurs publications : réalisation de scripts Python pour tester de nouvelles techniques de vectorisation et *clustering*, création de graphes avec Gephi, réorganisation du répertoire Github avec documentation

Stage de Recherche (M1 - 35h/semaine)

Juillet 2022 - Septembre 2022

LORLA, encadrée par Karěn Fort et Maxime Amblard

Nancy, France

- Réalisation d'un état de l'art sur les notions linguistiques de *hedging* et *claims*
- Constitution et pré-traitement d'un corpus d'articles ACL
- Extraction de *claims* en corpus, utilisation de méthodes de *clustering* en Python
- Évaluation de méthodes d'apprentissage non-supervisé

Stage de Recherche (L3 - 35h/semaine)

Juin 2021 - Août 2021

LORLA, encadrée par Karěn Fort, Gaël Lejeune et Yves Lepage

Nancy, France

- Annotation manuelle d'articles de TAL (TALN, LREC, ACL)
- Classification automatique de ces articles selon la présence de mention de la langue étudiée
- Travaux acceptés par la suite à LREC et TALN

Autres contributions

Chargée de TD de programmation Python

Université de Lorraine

Septembre 2023 -

Nancy, France

- Explications, aide et correction des TD d'étudiant-es en Master 1 de Traitement Automatique des Langues ou Master 1 de Sciences Cognitives.

Tutrice en informatique pour la Licence Sciences du Langage

Sorbonne Université

Septembre 2021 - Mars 2023

Paris, France

- Permanence de 3 heures par semaine auprès d'étudiant-es de la Licence Sciences du Langage pour aider à la réalisation des TD et projets de programmation Python, développement Web (HTML et CSS), ingénierie de la langue (TXM, Antconc, GATE), logique du premier ordre, ...

Participation à des comités de lecture

- Relectrice de deuxième niveau pour ACL 2023 (1 article)
- Relectrice de deuxième niveau pour TALN 2023 (1 article)

Participation à l'organisation de conférences

- Étudiante volontaire à ACL 2023 (gestion de session de posters, gestion du bureau d'enregistrement)
- Étudiante volontaire à IWCS 2023 (gestion des pauses cafés)

Publications

Conférences internationales avec comité de lecture

The Elephant in the Room: Analyzing the Presence of Big Tech in Natural Language Processing Research. Mohamed Abdalla, Jan Philip Wahle, Terry Ruas, Aurélie Névéol, Fanny Ducel, Saif Mohammad, Karën Fort. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, Juillet 2023, Toronto, Canada. - *CORE A**. <https://arxiv.org/pdf/2305.02797.pdf>

Participation : Réunions et discussions fréquentes avec 6 autres scientifiques de 3 pays différents prenant part au projet. Extraction, analyse et mise en forme de données brutes par scripts Python, annotations manuelles, relecture et correction de l'article.

Do we Name the Languages we Study? The #BenderRule in LREC and ACL articles. Fanny Ducel, Karën Fort, Gaël Lejeune, Yves Lepage. LREC 2022 - International Conference on Language Resources and Evaluation (LREC), Juin 2022, Marseille, France. - *CORE C, QUALIS B2*. <https://aclanthology.org/2022.lrec-1.60.pdf>

Participation : Présentation orale plénière à LREC, annotation manuelle d'articles de TAL, constitution et pré-traitement d'un corpus d'articles à partir de l'ACL Anthology, implémentation de classifieurs automatiques en Python, analyses et mise en forme des résultats, rédaction de l'article avec une approche éthique appelant à la diversité, réalisation des transparents pour la présentation.

Conférences nationales avec comité de lecture

Langues « par défaut » ? Analyse contrastive et diachronique des langues non citées dans les articles de TALN et d'ACL. Fanny Ducel, Karën Fort, Gaël Lejeune, Yves Lepage. Conférence sur le Traitement Automatique des Langues Naturelles (TALN), Juin 2022, Avignon, France. <https://aclanthology.org/2022.jeptalnrecital-taln.14.pdf>

Participation : Il s'agit d'un deuxième volet du travail présenté à LREC, les tâches sont donc très similaires, mais sur des données différentes. Réalisation et présentation du poster à TALN, et mêmes tâches que pour le travail LREC.

Workshops

Bias Identification in Language Models is Biased. Fanny Ducel, Aurélie Névéol, Karën Fort. Workshop on Algorithmic Injustice, Juin 2023, Amsterdam, Pays-Bas.

Participation : Réalisation d'un état de l'art sur les biais dans les modèles de langues avec annotation d'une soixantaine d'articles, extraction de méta-données et rédaction d'un résumé étendu critique de cet état de l'art pour un *workshop*.

Mémoires

Évaluation des biais stéréotypés dans les modèles de langue auto-régressifs : état de l'art et exploration de techniques extrinsèques Fanny Ducel.

https://fannyducel.github.io/_pages/MemoireM2Fanny.pdf

Analyse des *claims* dans les articles de Traitement Automatique des Langues à l'aide d'une méthode par apprentissage non supervisé. Fanny Ducel. 2022. <https://tinyurl.com/2p976fxf>.

Compétences informatiques

Langages : Python, SQL, SPARQL, Voice XML, UML, HTML, CSS, Java, C++, \LaTeX , Markdown

Outils de TAL : TXM, Antconc, Unitex, Praat, GREW-match, GATE, HTK, MBROLA, espeak, audeep

Autres outils : Neo4j, GitHub, Apache Solr, GANT, Agile (Scrum), Gephi, Twitter API

Librairies Python : Scikit-learn, SpaCy, NLTK, CoreNLP, Gensim, Pandas, LangChain, Pytorch

Autres intérêts

Membre de la communauté *Queer in AI*: Interactions et échanges de ressources pour une meilleure inclusivité et diversité dans le domaine de l'Intelligence Artificielle.

Trompettiste: Membre de l'École des Arts de la ville d'Orly pendant 8 ans.